EVOLUTION AND DYNAMIC BEHAVIOR OF TRANSFER RNA IN THE FIRST TWO STEPS OF TRANSLATION

BY

ALEXIS ANNE BLACK PYRKOSZ

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Chemistry
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2010

Urbana, Illinois

Doctoral Committee:

Professor Zaida A. Luthey-Schulten, Chair
Assistant Professor Douglas Mitchell
Assistant Professor Emad Tajkhorshid
Professor Wilfred A. van der Donk

ABSTRACT

In protein synthesis, a key component of the cellular machinery is transfer RNA (tRNA). This small nucleic acid is crucial to the maintenance of the genetic code because it discriminately binds the messenger RNA codon at the ribosome and adds the cognate amino acid to the growing polypeptide chain. The role of tRNA as an adaptor molecule has been understood for decades, but details about the charging of tRNA with cognate amino acids prior to entering the ribosome are still emerging. Aminoacyl-tRNA synthetases (aaRSs) are enzymes that recognize specific tRNAs and amino acids from the cellular pool and facilitate the charging of the correct amino acids on tRNAs. Following aminoacylation, tRNAs dissociate from the aaRSs and bind the elongation factor Tu (EF-Tu) for delivery to the ribosome.

The recognition of specific tRNA species by the aaRSs, EF-Tu, and other enzymes along the translation pathway is based on sets of highly conserved nucleotides within different groups of tRNA species. Previous work to identify these recognition elements has focused on experimental studies of single organisms. Here, bioinformatic analyses are used to predict recognition elements for groups of tRNA organized by domain of life and specificity. Shannon entropy differences between evolutionary profiles of tRNA domain/specificity groups and the representatives of all tRNA species reveal the uniquely conserved nucleotides within each tRNA domain/specificity, consistent with experiment. Comparative analysis of consensus sequences for these evolutionary profiles is used to locate tuning elements, also consistent with experiment. The discriminator base and the G53·C63 base pair are identified as conserved in several tRNA domain/specificities, particularly among *Archaea*. Both sets of predictions expand on the current knowledge of

recognition elements, providing suggestions for new mutation studies.

AaRS·tRNA complex formation and the aminoacylation reaction are well-characterized through many high resolution crystal structures and biochemical assays, but dissociation of the charged tRNA with subsequent binding to EF-Tu is not well understood. Using molecular modeling and molecular dynamics simulations, the effects of protonation states and the presence/absence of substrates and EF-Tu on tRNA release are explored. Using multiple dynamics and energetics analyses, the migration of protons from the $3'$ end of the tRNA and the α–ammonium group on the charging amino acid is shown to accelerate tRNA dissociation. The presence of AMP has only a minimal effect. Further, pKa calculations predict that Glu41, a conserved residue binding the α–ammonium group of the charging amino acid, is part of a proton relay system for releasing the charging amino acid upon transfer. This system is conserved both in structure and sequences across homologous aaRSs and may represent a universal handle for binding and releasing the charging amino acid. Addition of EF-Tu to the aaRS·tRNA complex stimulates tRNA dissociation. Knowledge of the exit strategies leads to a greater understanding of tRNA dynamics between the first two steps of translation.

To my husband and parents for their love and support.

ACKNOWLEDGMENTS

Behind every personal accomplishment are the people who provided support, encouragement, and mentoring. This work is the product of struggles and dreams from my six years of graduate education and I am deeply indebted to the following people.

To the faculty at UIUC:

Dr. Zaida Luthey-Schulten for taking a chance by allowing me to switch from experimental to computational research.

Dr. Wilfred van der Donk for initiating the NIH Chemical Biology Interface Training Grant program, providing support and constructive criticism for over five years, and acting as a secondary mentor.

Dr. Klaus Schulten for including members of the ZLS group in TCBG NIH Resource activities and for teaching that all science begins with one great question.

Dr. Susan Martinis for encouragement, constructive criticism, and up-to-date knowledge on all experimental tRNA issues.

Dr. Emad Tajkhorshid and Dr. Douglas Mitchell for joining my defense committee on short notice.

Dr. Steven Zimmerman for opening up opportunities in the department.

Dr. Anne Baranger for a collaboration wherein I realized my dream of performing computations on antibiotics.

Dr. Scott Silverman for encouragement at each departmental checkpoint.

Dr. Yi Lu for encouragement and constructive criticism during my literature seminar.

Dr. Paul Hergenrother for allowing me to spend a summer in his lab and get my bearings before my first year of graduate school began.

Dr. Virginia Leyman for support while teaching during my first year.

To the graduate students, postdocs, and staff at UIUC:

The ZLS group past and present who all contributed to my development as a scientist:

Dr. Rommie Amaro, Dr. Patrick O'Donoghue, Dr. Taras Pogorelov, John Eargle, Elijah Roberts, Damien Mathew, Leonardo Trabuco, Jonathan Montoya, Ke Chen, Li Li, Leonardo Sepulveda, Andrew Magis, and Shawn Wilkinson.

The Chemical Biology Interface Training Grant members and the fabulous manager Martha Freeland.

Members of the Schulten and Baranger groups, particularly Elizabeth Villa, John Stone, and Stacie Richardson.

Dr. Ellen Wang Althaus for advising the Women in Chemistry Committee and being a true friend.

The Women in Chemistry Committee members for social interaction and support.

The dedicated secretaries in the department and general chemistry offices for cheerfully handling all things administrative.

To the people who helped me before and outside of graduate school:

The faculty of Lawrence Technological University, my undergraduate alma mater, particularly Dr. William Madden, Dr. Walter Dean, Dr. Jerry Crist, Dr. Nicole Villenueve, Dr. Anthony Sky, Lavetta Appleby, Dr. Scott Schneider, and Dr. Melinda Weinstein.

Faculty mentors during my Research Experience for Undergraduates internships, Dr. Mark Hollingsworth and Dr. Donald Burke.

The musicians and runners of Champaign-Urbana.

The family and friends in Detroit.

My beloved husband who has supports me each day and keeps our long distance marriage together.

My parents and brother who taught me to strive, to seek, to find, and not to yield. I'm everything I am because you loved me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION TO TRANSFER RNA AND TRANSLATION

Transfer RNA (tRNA) is a small, ubiquitous RNA that provides the crucial link between nucleic acids and proteins in the cell. First discovered in the mid-twentieth century, tRNA has been at the forefront of the emerging field of molecular biology because of its direct influence on translation fidelity during protein synthesis (see [1] for a review). In Francis Crick's original hypothesis [2], tRNAs were adaptor molecules, reading the codons of the messenger RNA (mRNA) with a complementary triplet of nucleotides (anticodon). TRNAs were charged with amino acids according to their anticodons and these would be added to a growing polypeptide chain at the ribosome. The important function of tRNA is adding the correct amino acid to the nascent protein, thereby maintaining the genetic code. TRNAs therefore are subject to a unique set of physical constraints in that their structures must be sufficiently similar to bind enzymes during maturation and protein synthesis, but also different so as to be identified as specific to one amino acid during charging and decoding [3]. The diverse environments on Earth have provided strong evolutionary pressure, causing tRNAs and other translation enzymes in the *Bacteria*, *Archaea*, and *Eukarya* domains of life to further differentiate [4]. The following sections review the lifecycle of tRNA in the cell with an emphasis on the precise recognition of tRNA species by various enzymes that ultimately leads to accurate translation.

The primary function of tRNA is to maintain the genetic code, the direct correspondence between each of the 64 codons (nucleotide triplets) and the 20 standard amino acids (22 amino acids in those organisms containing pyrrolysine and selenocysteine [5]).

a) CODON TABLE

UUU	PHE	UCU	SER	UAU	TYR	UGU	CYS
UUC	PHE	UCC	SER	UAC	TYR	UGC	CYS
UUA	LEU	UCA	SER	UAA	STOP	UGA	STOP
UUG	LEU	UCG	SER	UAG	STOP	UGG	TRP
CUU	LEU	CCU	PRO	CAU	HIS	CGU	ARG
CUC	LEU	CCC	PRO	CAC	HIS	CGC	ARG
CUA	LEU	CCA	PRO	CAA	GLN	CGA	ARG
CUG	LEU	CCG	PRO	CAG	GLN	CGG	ARG
AUU	ILE	ACU	THR	AAU	ASN	AGU	SER
AUC	ILE	ACC	THR	AAC	ASN	AGC	SER
AUA	ILE	ACA	THR	AAA	LYS	AGA	ARG
AUG	MET	ACG	THR	AAG	LYS	AGG	ARG
GUU	VAL	GCU	ALA	GAU	ASP	GGU	GLY
GUC	VAL	GCC	ALA	GAC	ASP	GGC	GLY
GUA	VAL	GCA	ALA	GAA	GLU	GGA	GLY
GUG	VAL	GCG	ALA	GAG	GLU	GGG	GLY

b) ARCHAEAL ANTICODONS

GAA	PHE	GGA	SER	GUA	TYR	GCA	CYS
UAA	LEU	UGA	SER				
CAA	LEU	CGA*	SER			CCA*	TRP
GAG	LEU	GGG	PRO	GUG	HIS	GCG	ARG
UAG*	LEU	UGG	PRO	UUG	GLN	UCG*	ARG
CAG	LEU	CGG*	PRO	CUG*	GLN	CCG	ARG
GAU	ILE	GGU	THR	GUU*	ASN	GCU	SER
UAU*	ILE	UGU	THR	UUU*	LYS	UCU	ARG
CAU*	MET	CGU	THR	CUU*	LYS	CCU	ARG
GAC	VAL	GGC	ALA	GUC	ASP	GCC	GLY
UAC	VAL	UGC	ALA	UUC*	GLU	UCC*	GLY
CAC	VAL	CGC	ALA	CUC*	GLU	CCC	GLY

c) BACTERIAL ANTICODONS

GAA*	PHE	GGA	SER	GUA*	TYR	GCA	CYS
UAA*	LEU	UGA*	SER				
CAA*	LEU	CGA	SER			CCA*	TRP
AAG	LEU	AGG	PRO			ACG*	ARG
GAG	LEU	GGG	PRO	GUG*	HIS	GCG	ARG
UAG	LEU	UGG*	PRO	UUG*	GLN	UCG	ARG
CAG	LEU	CGG	PRO	CUG	GLN	CCG	ARG
		AGU	THR				
GAU	ILE	GGU	THR	GUU*	ASN	GCU	SER
UAU	ILE	UGU*	THR	UUU*	LYS	UCU*	ARG
CAU*	MET/ILE	CGU	THR	CUU	LYS	CCU	ARG
GAC	VAL	GGC	ALA	GUC*	ASP	GCC	GLY
UAC*	VAL	UGC*	ALA	UUC*	GLU	UCC*	GLY
CAC	VAL	CGC	ALA	CUC	GLU	CCC	GLY

d) EUKARYAL ANTICODONS

		AGA*	SER				
GAA*	PHE			GUA*	TYR	GCA	CYS
UAA*	LEU	UGA*	SER				
CAA*	LEU	CGA	SER			CCA*	TRP
AAG*	LEU	AGG*	PRO			ACG*	ARG
				GUG*	HIS		
UAG	LEU	UGG*	PRO	UUG*	GLN	UCG	ARG
CAG	LEU	CGG	PRO	CUG	GLN	CCG	ARG
AAU*	ILE	AGU*	THR				
GAU	ILE			GUU*	ASN	GCU	SER
UAU	ILE	UGU	THR	UUU*	LYS	UCU*	ARG
CAU*	MET	CGU	THR	CUU	LYS	CCU	ARG
AAC*	VAL	AGC*	ALA				
				GUC*	ASP	GCC	GLY
UAC*	VAL	UGC	ALA	UUC*	GLU	UCC*	GLY
CAC	VAL	CGC	ALA	CUC	GLU	CCC	GLY

Figure 1.1: Standard codon table and anticodon usage tables per domain of life. a) shows the standard codon table. b), c), and d) show the anticodons used for each specificity within each domain of life. Anticodons not observed in any organism within a given domain of life have been deleted from the table. Anticodons with a * have the wobble base modified in at least one organism according to [6].

Degenerate codons usually contain the same first and second nucleotide, but differ in the third. The third nucleotide is sometimes called the wobble base, which often can form noncanonical basepairs with the wobble base (first nucleotide) in the anticodon on the tRNA. As a result, a single tRNA may decode more than one codon; fewer than 64 anticodons are needed to decode mRNA (see [7] for a recent review). The mode of wobble base usage differs across tRNAs of varying amino acid specificity and domain of life with some organisms containing a single tRNA to decode up to three codons and some organisms containing five different tRNA isoacceptors with different anticodons that are all specific to the same amino acid (see Figure 1.1).

1.0.1 Tertiary structure

All canonical tRNAs contain between 74 and 95 nucleotides and fold into a specific secondary and tertiary structure (see Figure 1.2). The secondary shape is often called the cloverleaf containing the acceptor stem, D or GG arm, anticodon arm, and T or common arm. The acceptor stem and common/D arm coaxially stack to form one branch while the anticodon and GG/D arms stack to form the second branch. Together, these form an L-shaped tertiary structure with multiple interactions between the highly conserved nucleotides in the GG loop and common loop as well as the variable loop,

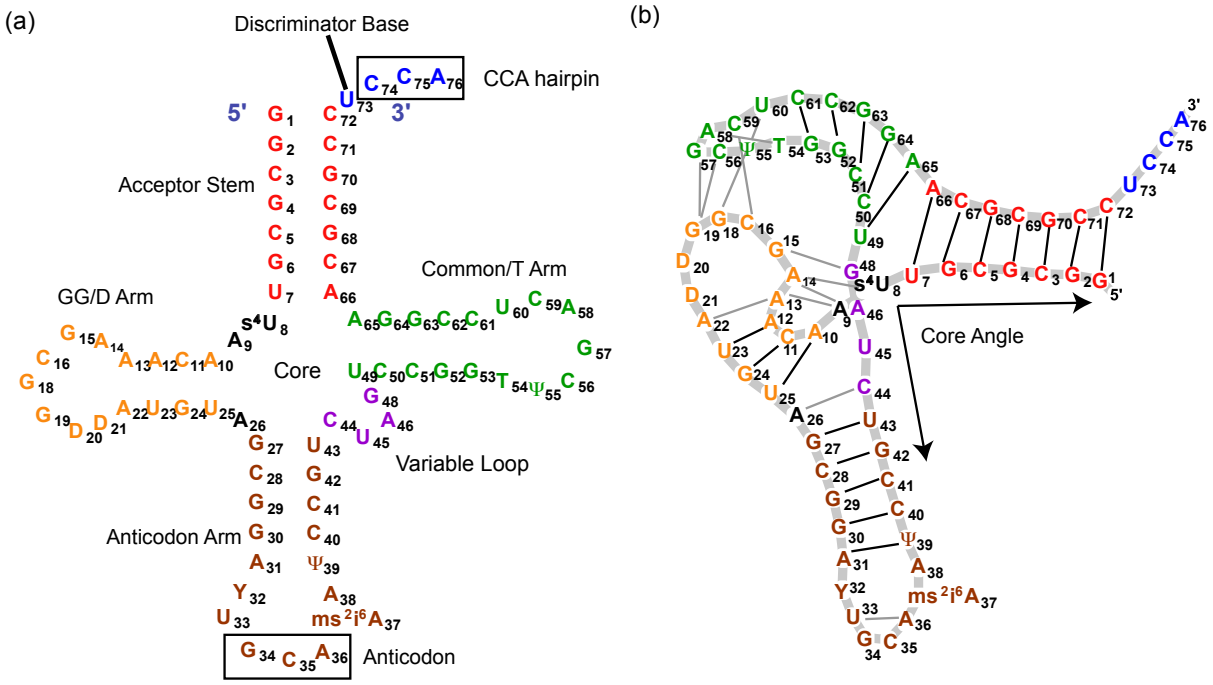


Figure 1.2: Example cloverleaf and L-shape schematics of *E. coli* tRNA^{Cys} with canonical numbering. Tertiary structure information was derived from the crystal structure [8]

forming a highly compact and globular core [9, 10]. The D/T nomenclature is characteristic of the bacterial/eukaryal tRNAs that often contain the modified nucleotides dihydrouridine and thymidine, respectively. Archaeal tRNAs rarely contain either and therefore the GG/common arm terminology is often used when making comparisons across all three domains of life. The L-shape structure imparts considerable flexibility to tRNA

with large variations in the angle at the tRNA core (junction of the two branches) occurring when tRNA undergoes induced conformational changes while binding to various translation enzymes (see [11] for a review).

1.1 TRNA maturation

Transcription of tRNA genes generates long nucleic acids that need to be processed by several different enzymes to become mature, functional tRNA. The first step is to trim the 5' end of the precursor tRNA (pre-tRNA) with RNase P and the 3' end with endonucleases that vary across domains of life [12, 13, 14]. The subsequent steps vary based on tRNA specificity and domain of life. One of the most common is the addition of modifications to various nucleotides (more will be said on this later). These modifications can aid in the folding of the nucleic acid to its final tertiary structure and stabilization.

1.1.1 Splicing and Editing

A step that is required in many archaeal and eukaryal organisms is the excision of introns, short nonfunctional segments of RNA embedded in the tRNA gene. Introns are usually located between nucleotides 37 and 38 (according to tRNA canonical numbering) almost immediately following the anticodon. The strategic positioning of the introns is recognized by endonucleases that splice the two halves of the tRNA together [15]. Bacterial tRNAs contain self-splicing group I introns. Archaeal organisms contain the largest number of intron insertions per tRNA gene, both in frequency of occurrence over all genes and number of introns within a given gene [16, 17, 18].

Additional steps may be required for processing the disrupted tRNA genes that often occur in archaeal and some eukaryal organisms. One is the splicing of split tRNAs. *Nanoarcheum equitans* was discovered in 2005 to contain six essential tRNA isoacceptor split genes with the halves associating following transcript excision [19, 20]. A 5' leader sequence on each 3' half matched the 3' trailing sequence on each 5' half, forming a clamp

between nucleotides 37 and 38 that would hold the two halves together until splicing could occur. With the more recent sequencing of the crenarcheon *Caldivirga maquilingensis*, $tRNA^{Gly}$ isoacceptor genes in three pieces were found with trailing/leader sequences forming similar clamps [21].

A step of minor sequence editing exists in some archaeal organisms. C-to-U editing was recently found in *Methanopyrus kandleri* where C8 in tRNA genes would be deaminated to U8 [22]. U8 is nearly universally conserved. The disparity between the gene sequences and mature tRNA sequences is evidence that hypotheses based solely on genomic information may be slightly inaccurate when applied to mature tRNA sequences.

One more unusual step of note is the addition of an extra guanosine to the 5' end of $tRNA^{His}$ in archaeal and eukaryal organisms. In bacteria, the trimming of the 5' end by RNase P results in $tRNA^{His}$ having G-1 that base pairs with C73 (position 73 is usually the last nucleotide on the 3' end and often called the discriminator base). When G-1 is not present in the gene (as in *Archaea/Eukarya*), $tRNA^{His}$-guanylyl-transferase adds it [23, 24, 25].

The final step of tRNA maturation is the addition of the CCA sequence motif to the 3' end. While some organisms contain the CCA in the gene, most use a ATP(CTP)·tRNA nucleotidyl-transferase to discriminately convert CTP and ATP substrates to the universal C74, C75, and A76 [26, 27].

1.1.2 Modified bases

All characterized tRNA species contain multiple modifications; one study estimated that a median of 8 modifications are present per tRNA [6]. While some modifications are simple (methylations of the bases or sugars or conversion of uracil to dihydrouracil), many are elaborate with large substituents added to bases or excision of a standard base with subsequent replacement by a nonstandard one (see [28] for a recent review). Over 100 chemically distinct modifications have been identified with some restricted to a particular domain of life, but most shared among all three as well as with organellular tRNAs [29].

The role of modified nucleotides is dual-purpose with one set of modifications affecting the binding of tRNA to various interaction partners and the other set influencing the folding and stability of the tertiary structure. Of the first set, the most significant modifications are in the anticodon loop (see [7] and references therein). Modification of the wobble base (position 34) can enable a single tRNA isoacceptor to decode multiple mRNA codons. Of the mature tRNA sequences known, U34 is modified in nearly every occurrence as is A34 [6]. Previous studies have hypothesized that modification of the wobble base can increase the rate of translation (by decreasing the number of precise matches between codon and anticodon) while still maintaining decoding accuracy (the first two bases in the codon must base pair with the second two bases in the anticodon) [30].

One specific example of wobble base modification is lysidine, a cytosine base with a lysine substituent. $tRNA^{Ile}$ in bacteria contains a CAU anticodon to bind the AUA codon (CAU is more commonly used as the anticodon in $tRNA^{Met}$) [31, 32]. In *E. coli*, this $tRNA^{Ile}$ isoacceptor with the lysidine modification is charged exclusively with isoleucine and successfully decodes only the AUA codon. The enzyme performing the modification, TilS, uses the acceptor stem (C4·G69 and C5·G68) and a modified A37 as recognition elements for binding $tRNA^{Ile}$(CAU) and not modifying $tRNA^{Met}$(CAU) (which contains U4·A69 and A5·U68, which are considered negative determinants for TilS recognition) [33].

Modifications that affect tRNA structural stability are mostly located in the tRNA core. The ubiquitous TΨC sequence motif in the common/T arm is present in all bacterial/eukaryal tRNAs (archaeal tRNAs contain other modifications at positions 54, 55, 56) and has been shown by temperature melting experiments to impart greater stability to tRNA structure. This is particularly seen at low Mg^{2+} concentrations (a melting temperature increase of nearly 6 degrees Celsius in modified versus unmodified tRNA) [34, 35, 36, 37, 38, 39, 40, 41, 42]. Pseudouridine and 2'O methylation of the ribose have also have stabilizing effects on helices [43, 44, 45, 46, 47, 48, 49]. However, not all modifications lead to greater structural rigidity. Dihydrouridine, present in the D arm of most bacterial/eukaryal tRNAs, imparts greater flexibility to the tRNA structure [50].

Interestingly, most archaeal organisms inhabiting environments with high temperatures (greater than 80 degrees Celsius) are devoid of dihydrouridine.

1.1.3 Cations

The L-shaped tRNA structure has one drawback; the elbow region contains many closely packed, negatively charged phosphate groups. Stabilization of this region is achieved by association of the phosphates with metal cations. Crystal structures of tRNAs have revealed that between 5 and 11 divalent ions are present, occupying specific sites in RNA grooves to stabilize the tertiary structure [51, 52, 53, 54]. Computational studies have shown that most ions near the tRNA are diffuse, providing nonspecific electrostatic interaction, with only a few ions directly bound. Mg^{2+} or related divalent ions are more likely to be present than monovalent ions because they reduce the electrostatic repulsion between phosphates on opposite sides of a RNA groove at lower entropic cost (fewer ions bound to tRNA) [55, 56], thereby adding stability. Divalent ions are present in many recent high resolution X-ray crystal structures of tRNA [57, 58, 59, 60, 61, 62] and molecular dynamics simulations have demonstrated that at least three Mg^{2+} ions are resident in the tRNA core[63].

1.2 Step 1: Aminoacylation

Mature tRNAs are charged with their cognate amino acids by the aminoacyl-tRNA synthetases (aaRSs), which recognize the correct amino acid and tRNA from the cellular pool. Aminoacylation is a two-step process. In the first step, ATP and the amino acid are converted to an aminoacyl-adenylate with subsequent formation of pyrophosphate. The activated amino acid is then transferred onto the $3'$ end of the tRNA. Many decades have been devoted to studying the details of this part of the translation pathway because the charging of the tRNA with the correct amino acid is crucial for maintaining the genetic code. The tRNA·aaRS pairs have coevolved to form highly specific complexes to ensure

correct charging; some aaRSs have also evolved additional proofreading mechanisms to edit misacylated tRNAs. The study of this portion of the pathway is hindered by the divergence of aaRSs; most organisms contain 20 different enzymes that differ across the three domains of life [64]. Despite over 200 high resolution X-ray crystal structure currently solved, a complete set of aaRS structures for a single organism is not available. Further, most structures are from bacterial organisms, and few structures contain tRNA [11].

1.2.1 Identity elements

Recognition of specific tRNAs by the corresponding aaRSs is mediated by identity elements, highly conserved nucleotides that collectively form a unique tRNA shape (see [65, 11, 66] for a review). Recognition ensures that tRNA is charged with the cognate amino acid. Identity rules are sometimes referred to as a second genetic code and consist of several determinants (positive recognition elements) and antideterminants (negative recognition elements). The fundamental characteristic of an identity element is purely experimentally derived; mutation of the nucleotide decreases the rate of aminoacylation [65]. As a result, early versions of the list of identity elements were mostly limited to those organisms easily cultured under standard laboratory conditions, particularly *Escherichia coli* and *Saccharomyces cerevisiae*. More recent studies have considerably expanded the diversity of organisms on the list [67], but there are still tRNAs for particular amino acids in the archaeal or eukaryal domains of life that have no known identity elements. Further, because some tRNAs coevolve with their respective aaRSs, identity elements specific to one tRNA species in a given organism may not be present in the same tRNA species in another organism. Most identity elements are located in a few specific tRNA regions: the anticodon arm (particularly the anticodon), the acceptor stem, and the GG/D arm.

1.2.2 Aminoacyl-tRNA synthetases

AaRSs consist of two primary domains: the catalytic domain that contains the active site and the anticodon binding domain that generally binds the anticodon loop of the tRNA. Based on the architecture of the catalytic domain, the 20 aaRSs are divided into two classes [68]. Class I aaRSs have a Rossmann nucleotide binding fold while class II aaRSs contain a 7-stranded β–sheet with flanking α–helices [69]. These two architectures appear at the base of the universal phylogenetic tree and are almost completely fixed (i.e. an aaRS that is Class I or II in one organism is the same class in all other known organisms). Each class is divided into three subgroups based on the homology of the anticodon binding domain (see [70] for a review). The anticodon binding domains evolved later and vary between aaRSs in different domains of life. The standard class assignment is: Class Ia–IleRS, LeuRS, MetRS, ValRS, ArgRS, and CysRS, Class Ib–GluRS and GlnRS, Class Ic–TrpRS and TyrRS, Class IIa–ProRS, SerRS, ThrRS, AlaRS, HisRS, Class IIb–AspRS, AsnRS, and LysRS, and Class IIc–PheRS, GlyRS.

The recognition and binding of tRNA is generally mediated by contacts between the anticodon and anticodon binding domain followed by the acceptor stem/CCA hairpin and catalytic domain [68]. The difference in binding between the two classes results in differing conformations of the CCA hairpin in the active site. In class I aaRSs, the CCA hairpin is contorted to fit into the active site [71], and the rate-determining step for aminoacylation is the dissociation of the charged tRNA [72, 73, 74, 75, 76]. Recent experimental studies have suggested that the elongation factor Tu (EF-Tu) may facilitate a faster release by binding the aaRS·tRNA complex before the tRNA has completely dissociated from the aaRS [77, 78]. In Class II aaRSs, the rate-determining step occurs prior to the transfer of the charging amino acid onto the tRNA [76], and is believed to be the amino acid activation step [79, 80, 81, 82].

With the fundamental docking of tRNA to aaRS established, several aaRSs have evolved additional domains to assist in maintaining aminoacylation fidelity and efficiency. AaRSs in higher organisms have larger aaRSs with insertions in both primary domains or

additional tRNA binding domains [83, 84], and can form homopolymers or even multi-aaRS complexes (see [85, 86] for reviews). Further, because misacylation can have potentially dire effects – cell death in single cell organisms and neurological disease in higher organisms [87, 88] – many aaRs have evolved proofreading mechanisms.

Editing

Previous studies have revealed that nearly half of all aaRSs contain a separate domain with a hydrolytic active site for editing tRNA [89]. Most of these aaRSs are specific for amino acids that are hydrophobic or otherwise difficult to distinguish from other amino acids. Editing domains are most commonly an insertion in the catalytic domain that bind the tRNA following aminoacylation [5]. Post-transfer editing occurs by recognizing the amino acid charged on A76 and hydrolyzing the wrong amino acids.

Indirect pathways

Not all organisms contain a complete set of 20 aaRSs. Asparagine, glutamine, and cysteine are known to be charged on their cognate tRNAs via an indirect mechanism in several organisms, particularly archeons. In these pathways, the correct tRNA is misacylated by a nondiscriminating aaRS (e.g. GluRS aminoacylates both $tRNA^{Glu}$ and $tRNA^{Gln}$ with glutamate). The incorrect amino acid is a precursor for the final amino acid and is converted by an additional enzyme. In the case of asparagine and glutamine, $tRNA^{Asn}$ and $tRNA^{Gln}$ are misacylated by AspRS and GluRS respectively and amidotransferases reduce the incorrect amino acids to the cognate ones (see [90] for a review). For cysteine, $tRNA^{Cys}$ is aminoacylated with the precursor *O*–phosphoserine by SepRS with subsequent conversion to cysteine by SepCysS, a pyroxidal phosphate dependent enzyme [91].

1.3 Step 2: Transportation to Ribosome

To protect the delicate ester linkage between the charging amino acid and A76, the elongation factor Tu (EF-Tu) or 1A (bacteria or archaea/eukarya respectively) binds all charged tRNAs and shuttles them to the ribosome. EF-Tu is a bacterial G protein that is highly conserved across all domains of life [63]. The GTP bound form binds all charged tRNAs and docks to the ribosome in a multistep mechanism [92, 93, 94, 95, 96, 97]. Consisting of approximately 400 residues, EF-Tu has three domains, a N-terminal GTP binding domain, and two β–barrel motifs. This step is another checkpoint for establishing fidelity for two reasons. First, the EF-Tu can recycle a prematurely released misacylated tRNA back to the aaRS for proofreading [98]. Second, it recognizes the tRNA specificity through conserved nucleotides at the docking interface.

1.3.1 Tuning elements

The constraints on tRNA to be both unique yet similar modulate the binding of tRNA to EF-Tu. Co-crystal structures of EF-Tu and tRNA indicate that the enzyme interacts both with the charging amino acid and the acceptor stem/common arm [99, 100]. Experimental studies have shown that the EF-Tu affinity for the tRNA body balances the affinity for the charging amino acid [101, 102], ensuring sufficiently strong binding for transport to the ribosome with weak enough binding to dissociate upon arrival. The presence of tuning elements, highly conserved nucleotides that form strong contacts to EF-Tu, balances weak binding amino acids.

1.4 Step 3: Decoding the message

Ribosomes are the translation machinery responsible for protein synthesis in all cells. They are comprised of two parts. The 30S or small subunit consists of the 16S rRNA and about 20 proteins (in bacteria). The 50S or large subunit contains the 23S rRNA, 5S rRNA and

approximately 30 proteins. Recent high resolution X-ray crystal structures of bacterial ribosome complexes have revealed the interactions between tRNA and the rRNAs in atomistic detail [97, 103, 104].

Protein synthesis is a four step process: initiation, elongation, termination, and recycling (see [105, 106] for recent reviews). During initiation, the subunits form a complex with the mRNA bound to the small subunit. The exact positioning is governed by base pairing interactions between the Shine-Dalgarno sequence at the beginning of the mRNA and the complementary anti-Shine-Dalgarno sequence at the 3' end of the 16S rRNA. An initiator tRNA is guided by initiation factors to bind the mRNA start codon, triggering the rapid association of the 50S subunit to the 30S, forming the full 70S ribosome. The three tRNA binding sites in the 70S ribosome are the aminoacyl (A), peptidyl (P), and exit (E) sites, which are located between the two subunits; the tRNA anticodon arm binds the mRNA cradled by the small subunit while the CCA hairpin and charging amino acid interact with nucleotides in the peptidyl-transferase center (PTC) in the large subunit [107, 108, 109].

Once the ribosome is assembled and an initiator tRNA is in the P site, the elongation step begins. EF-Tu·tRNA·GTP ternary complexes interact with the A site where the tRNA anticodon attempts to bind the mRNA codon while the acceptor stem remains bound to the EF-Tu [110]. This is known as the A/T state and only correct matches between the anticodon and codon form a stable state [111]. Noncognate and near cognate complexes form unstable A/T states and are rejected [112]. A stable A/T state triggers hydrolysis of GTP on the EF-Tu, causing the charged or aminoacyl-tRNA to dissociate from EF-Tu with the acceptor stem and CCA hairpin being accommodated into the 50S part of the A site. The charging amino acid on the A site tRNA is added to the growing peptide in the PTC and the tRNA is shifted to the P site. Entrance of the next cognate tRNA in the A site and addition of the next amino acid in the growing peptide causes the chain to shift from the P site tRNA to the A site tRNA, resulting in a deacylated tRNA in the P site. Finally, the deacylated P site tRNA is translocated to the E site, a reaction that is catalyzed by elongation factor G hydrolyzing GTP [113]. From the E site, the tRNA exits the ribosome

to either be aminoacylated by the aaRS again or undergoes degradation. When the mRNA stop codon is reached, protein synthesis ends (termination). The mRNA, tRNA, and other protein factors dissociate from the ribosome, and the two subunits split apart (recycling).

1.4.1 Recognition elements

While experimental studies have confirmed that conserved nucleotides in the 16S rRNA discriminate between cognate and noncognate tRNAs by interacting with the codon-anticodon helix [114], the greater question is whether tRNAs contain recognition elements that also assist in uniformly binding to the ribosome. Common questions are whether a GC rich codon-anticodon pair will form a tighter helix faster than an AU-rich pair or whether the identity of the amino acid will slow movement through the ribosome (small nonpolar amino acid versus bulky aromatic amino acid). Recent studies have demonstrated that tRNAs move through the ribosome at a uniform rate despite these factors [115, 116]. Structural data suggest that extensive contacts between aminoacyl-tRNA and ribosome are formed at each step of protein synthesis [114, 111, 104], indicating that the mode of binding differs between varying tRNA species. While it is possible that the ribosome accommodates these differences through conformational changes, it has been suggested that tRNAs are tuned for ribosome interaction similar to elongation factor tuning. Mutation studies have demonstrated that changing portions of the tRNA dramatically decreases their ability to function during decoding [117, 118, 119, 120, 121, 122, 123]. Further studies have identified several modified nucleotides that are also essential for accurate and efficient decoding [124, 125, 126, 127, 128]. A few studies have attempted to identify the exact recognition elements needed to achieve this uniformity of binding during decoding [129, 130], but more work is needed in this area.

1.5 Summary

Accurate recognition of specific tRNA species at each step of the translation pathway is of crucial importance for protein synthesis. During tRNA maturation, the tRNA must be correctly folded, modified, and spliced (if necessary) to form functional molecules that have a similar structure but distinctly different features. In the first step of translation, tRNA is recognized by the aaRS through identity elements, allowing tight binding during aminoacylation. Still unknown is the dissociation mechanism for charged tRNA from the aaRS prior to binding EF-Tu, an issue that will be explored in Chapters 3 and 4. During association of aminoacylated tRNA with EF-Tu, tuning elements are used to screen misacylated tRNAs prior to reaching the ribosome. At the ribosome itself, recognition elements in addition to the anticodon are used to screen tRNAs prior to accommodation in the A site. Locating these numerous recognition elements (including identity and tuning) will be addressed on a large-scale in Chapter 2. Together, the following chapters add to the knowledge base of tRNA fidelity and dynamics at multiple steps along the translation pathway.

CHAPTER 2

DETERMINATION OF IDENTITY/TUNING/RECOGNITION ELEMENTS IN TRNA

2.1 Summary

The goal of this study is to predict those uniquely conserved nucleotides within domains of life/amino acid specificities that serve as recognition sites for tRNA binding to translation enzymes, thus ensuring amino acid fidelity at the ribosome. An extensive bioinformatics studies is performed, involving the alignment of over 50,000 tRNA sequences from the maximum diversity of organisms. Using Shannon information entropy, the uniquely conserved nucleotides in each tRNA domain of life/amino acid specificity are determined. These nucleotides show good agreement with identity/tuning/recognition elements previously determined by experiment. Those uniquely conserved nucleotides that have no experimental equivalent are predicted to be important recognition elements for future studies. Also, consensus sequence comparisons are used to locate tRNA domain/specificities that contain tuning elements [1].

[1]Data from this work was used in [63, 131, 132, 133]. Portions of text and figures were previously published in John Eargle, Alexis Black, Anurag Sethi, Leonardo Trabuco, and Zaida A. Luthey-Schulten "Dynamics of Recognition between tRNA and Elongation Factor Tu" *J. Mol. Biol.* 2008, 377(5), 1382–1405 © 2008 Elsevier and Rebecca W. Alexander, John Eargle, Zaida Luthey-Schulten. "Experimental and computational determination of tRNA dynamics" *FEBS Lett.* 2010, 584, 376–386 © 2010 Elsevier (ABP supplied a figure). Design of the study, data validation and processing, consensus sequence analysis, and entropy difference analysis were performed by ABP. Scripting assistance was provided by Dan Wright and John Eargle. Use of entropy differences for identifying tuning elements was first suggested by Anurag Sethi.

2.2 Introduction

Transfer RNA (tRNA) is crucial to translation in the cell. TRNAs dock aminoacyl-tRNA synthetases (aaRSs), which charge tRNAs with the correct amino acid according to tRNA specificity (see [68] for a review). TRNA is recognized by the cognate aaRS via highly conserved nucleotides unique to a particular amino acid specificity (identity elements). To protect the delicate ester linkage of the charged amino acid to the adenosine at the $3'$ end, an elongation factor (EF-Tu in *Bacteria* and EF1A in *Archaea* and *Eucarya*) shuttles the tRNA to the A-site in the ribosome ([134]). To facilitate docking and dissociation from the EF-Tu, tRNAs also contain tuning elements, nucleotides that stabilize the complex if the charged amino acid binds weakly in the EF-Tu active site. At the ribosome, the anticodon binds to the complementary base triplet of the messenger RNA (mRNA), and the amino acid is added to the growing polypeptide chain. Decoding of the mRNA by the tRNA sets the genetic code. While identity/tuning/recognition elements are known to be essential for cognate tRNA binding to various enzymes along the translational pathway, the particular nucleotides that participate in docking are often unknown. Many experimentalists have mutated nucleotides to find the set that has the greatest effect on binding (see [65, 67] for reviews). For practical reasons, these studies have been limited to only a few organisms. Further, these studies focus on a single step of the translation. The question is whether these identity/tuning/recognition elements can be predicted computationally, thereby removing the need to perform extensive mutation studies. The goal of this study is to use bioinformatic methods to predict identity/tuning/recognition elements for all 20 tRNA amino acid specificities across three domains of life. Experimental data support many of these predictions, and the unconfirmed elements suggest regions for further study.

2.2.1 Availability of diverse tRNA data

Compilations of tRNA sequences have existed for more than 40 years [6]. Initially scarce, tRNA sequences are now readily availabie due to recent genome sequencing initiatives.

TRNA sequences are identified within genomes [135, 136] and annotated by amino acid specificity. For the current study, two compilations were sourced. The first was the tRNA Compilation 2000 [6], which has a long history of reliable curation and contains alignments of nearly 2,000 tRNAs. The second was the Integrated Microbial Genome Database [136], which contains over 4,000 genomes from recent sequencing projects, with most genomes containing a full set of tRNA.

One reason for the lack of up-to-date databases of tRNA alignments is because alignment of tRNA sequences is a time consuming process requiring manual curation. Most modern tRNAs exhibit a well-defined structure, but contain two highly variable regions (see Figure 2.1). In the variable loop, different tRNA specificities and domains of life have varying lengths and base pairing must be manually curated for those with a long stem.

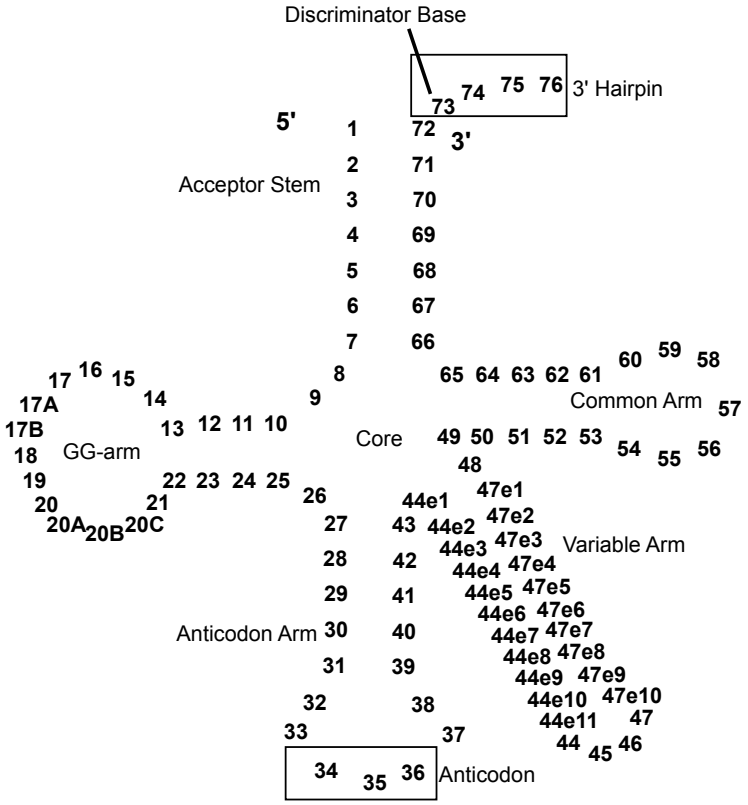


Figure 2.1: Cloverleaf schematic of tRNA with canonical numbering. GG-arm and common arm terminology appropriate to tRNAs in all domains of life have been used to label the different parts. Alphanumeric labels denote positions where nucleotides are rarely present and are concentrated in the GG-arm and variable arm

Also, some tRNAs identified from whole genome sequencing are pseudogenes, ancient genes that are no longer being used and are degraded, which can be difficult to automatically screen. For this study, the tRNA sequences from the selected data sources were aligned and manually curated. The current alignment contains over 50,000 tRNA sequences, the largest and most diverse set currently known.

Species bias is an inherent problem in any large data set of biological sequences or database [137, 138]. Due to limitations on diversity sampling or slowly evolving taxa, over-representation of particular groups of organisms can prejudice bioinformatics studies and lead to conclusions that are not applicable to the full range of phylogenetic diversity. The SeqQR algorithm uses a Householder Transformation to calculate the most linearly independent set of sequences from a multiple sequence alignment [139], thereby reducing a profile to the minimum number of sequences required to represent the diversity of the full set. This algorithm has been used effectively in several bioinformatics studies [139, 63, 131, 132, 140, 141] and is used in this study to create well-balanced evolutionary profiles within groups of domain of life/amino acid specific tRNAs (domain/specificity).

2.2.2 Shannon entropy and consensus sequences as measures of unique conservation

The uniqueness of conserved nucleotides within tRNA domain/specificities can be measured using differences in Shannon entropy [142, 143] and consensus sequences [144]. Shannon entropy, also known as information theory, is a measure of uncertainty. Given a multiple sequence alignment and the occurrence of each nucleotide in a given column, information is knowledge of the identity of the next nucleotide in that column. Nucleotides that commonly occur in an alignment column (conserved) are considered to have low information while those that are rare (unconserved) have high information content. By taking the difference between entropies for an alignment containing representatives of all tRNAs and an alignment with a specific group of tRNAs, those nucleotides that are more

conserved in the group than the full set are shown to have large, positive values. In this study, the high information nucleotides are shown to correspond with known identity/tuning/recognition elements in many groups of tRNA. Setting a threshold for information allows for prediction of other high information nucleotides to be additional identity/recognition elements. Tuning elements are shown to be more difficult to predict from the entropy difference analysis due to weaker signals and therefore are analyzed using consensus sequences over multiple groups of tRNA.

2.3 Methods

Data sources

The tRNA Compilation 2000 sequence database [6][2] was taken as the core alignment; all other sequence data was downloaded from the Integrated Microbial Genome database at the Joint Genome Institute [145][3] [4]. Sequences from both final and draft genomes were used and grouped by specificity according to gene annotation. Organellular and split tRNA gene sequences, as well as mature sequences with modified bases were not considered. In the final dataset, 143 archaeal, 629 bacterial, 387 eukaryal sequences from the tRNA Compilation 2000 database and 1,705 archaeal, 48,763 bacterial, and 1,708 eukaryal sequences from the IMG database were used. This resulted in a total of 53,335 sequences. The disparity in number of sequences between domains of life is characteristic of the wealth of information available for bacteria with whole genome sequencing of archaeal and eukaryal organisms progressing more slowly.

[2](http://www.staff.uni-bayreuth.de/ btc914/search/index.html)

[3]http://img.jgi.doe.gov/cgi-bin/pub/main.cgi

[4]At the time of this study, these were the two most complete and accurate repositories. Additional databases have become available or increased in accuracy in recent years [146, 147, 148]

Development of tRNA profiles

The sequences were processed separately depending on origin because the well-curated tRNA Compilation 2000 sequences were aligned while the IMG sequences required alignment and splicing. The IMG sequences were divided by specificity and domain of life, and aligned with ClustalW [149] with manual improvement in MultiSeq [150] in VMD [151]. Introns were excised by comparison with the tRNA Compilation 2000 mature tRNA sequences. Because the IMG data contained draft genomes from newly sequenced environmental samples, it was assessed for pseudogenes and sequencing errors by comparing poorly aligned sequences to the rest of the group. Sequences with longer ends were truncated. Sequences with shorter ends, insertions or deletions in uniform regions (variable regions are noted in Figure 2.1 with A, B, e, etc), incorrect anticodon for the specificity (except when known to be modified to the correct anticodon as reviewed in Chapter 1), characters other than the standard nucleotides, or acceptor stems that could not form Watson-Crick or common noncanonical base pairs were removed from the dataset. Data from the tRNA Compilation 2000 was similarly screened. The full alignments were deposited in an public database at http://trna.scs.uiuc.edu/RNADB.

To create a master evolutionary profile of aligned tRNA across all specificities and domains of life, each alignment was fit to the basic profile format. This format (Figure 2.1) with variable length regions was developed by using the tRNA Compilation 2000 profile format with additional gaps accounting for the variable regions. Table 2.1 shows the number of sequences in each subgroup of the master evolutionary profile.

The program SeqQR [5] was used to create evolutionary profiles (parameters: norm order = 3 and gap penalty = 0.5). These profiles have a maximum of 85% sequence identity, and are therefore subsets that are representative of the diversity displayed in all species within a given set in a sufficiently small group for further analysis.

The information entropy for each sequence profile is used as a measure to distinguish between nucleotides that are highly conserved within particular specificities and those that

[5] http://www.scs.uiuc.edu/∼schulten/software/index.html

AA	Archaea		Bacteria		Eukarya	
	Full	QR85	Full	QR85	Full	QR85
Ala	158	19	3337	41	167	25
Arg	186	37	4212	64	165	43
Asn	53	12	1692	42	64	13
Asp	58	8	1977	43	81	8
Cys	47	17	1137	35	41	11
Gln	77	19	1826	63	95	21
Glu	83	11	2254	56	140	17
Gly	191	22	3566	47	176	23
His	42	11	1042	29	51	12
Ile	53	9	1964	43	115	14
Leu	226	35	5012	71	107	33
Lys	91	12	2571	46	165	17
Met	121	23	3400	63	68	12
Phe	59	13	1289	38	77	9
Pro	73	19	2365	55	108	15
Ser	182	43	3544	76	156	27
Thr	140	21	2869	56	128	27
Trp	29	11	999	48	42	11
Tyr	36	9	1337	48	58	9
Val	138	22	2999	60	145	20

Table 2.1: Number of sequences in full alignments and evolutionary profiles per specificity and domain of life.

are universally conserved. The entropy (S) for all tRNAs was calculated across the master alignment for each column i; $S_i = -\sum_{j=1}^{4} P_{ij} \log_2 P_{ij}$. The entropy was then calculated for each tRNA domain/specificity, combining tRNA species for a given amino specificity within a domain of life into a domain/specificity alignment. Note that for both the master and domain/specificity alignments, the entropy is only calculated for columns with $\geq 20\%$ non-gap rows, preventing false positives in the variable regions. Using the formula stated above, this resulted in the sequence entropy for each domain/specificity, $S_{i|dAA}$. Next, the difference was taken between these entropies: $\Delta S_i = S_{i|dAA} - S_i$. Those nucleotides in each domain/specificity that are uniquely conserved are distinguished by their high ΔS_i. These elements have a high information content relative to the master profile, which forms the basis for using them to predict the identity/tuning/recognition elements.

The consensus sequences were generated according to the algorithm developed previously [144]. The probability P_{ij} is a measure of occurrence of a particular nucleotide j in a given column i; $P_{ij} = \frac{N_{ij}}{N_{seq}}$. This is taken over the number of sequences N_{ij} that contain a nucleotide in column i over the total number of sequences N_{seq}. A nucleotide is deemed the consensus nucleotide for a given column if $P_{ij} > 0.5$ and $P_{ij} > 2 \cdot P_{ik}$, where P_{ik} is the probability of occurrence for the second most common nucleotide in that column. If these two conditions are not satisfied but $P_{ij} + P_{ik} > 0.75$, then the nucleotides are co-consensus nucleotides. The common expanded nucleotide single letter code has been used with R and Y indicating purine (A or G) and pyrimidine (C or U), S and W for strong (G or C) and weak (A or U), and K and M for ketone (G or U) and amino (A or C).

2.4 Results and Discussion

2.4.1 Confirmation of identity elements

To assess whether the entropy difference analysis is successful in determining identity/tuning/recognition elements, the well-studied case of $tRNA^{Ala}$ was used to compare the computational results with previous experimentally determined elements. As seen in Figure 2.2, the identity elements previously identified correspond to high information nucleotides (large positive entropy difference). Base pairs G2·C71 and G3·U70 in the acceptor stem are both uniquely conserved in bacteria, which agrees with experiment [119, 152]. A73 (the discriminator base) is a high information nucleotide and an identity element in $tRNA^{Ala}$. The nucleotide at position 20 is known to be an identity element from experiment, but is shown here to not be high information nucleotide. It is possible that this identity element is specific to a subgroup of organisms. Also of note is the negative entropy difference seen in positions 32 and 38. In nearly all tRNAs, position 32 contains cytidine or uridine modified to have cytidine-like characteristics while position 38 is adenine. In all three bacterial $tRNA^{Ala}$ isoacceptors, these positions contain a mixture of nucleotides, most frequently an adenine and uridine. It was recently reported that

positions 32 and 38 in tRNA^{Ala} are uniquely important for cognate codon reading at the ribosome [130], suggesting that nucleotides with negative high information can also be important recognition elements. Finally, several nucleotides have not been identified as important by experiment, but are shown to be uniquely conserved by the entropy difference analysis. These are possible targets for future mutation studies and will be discussed in the next section.

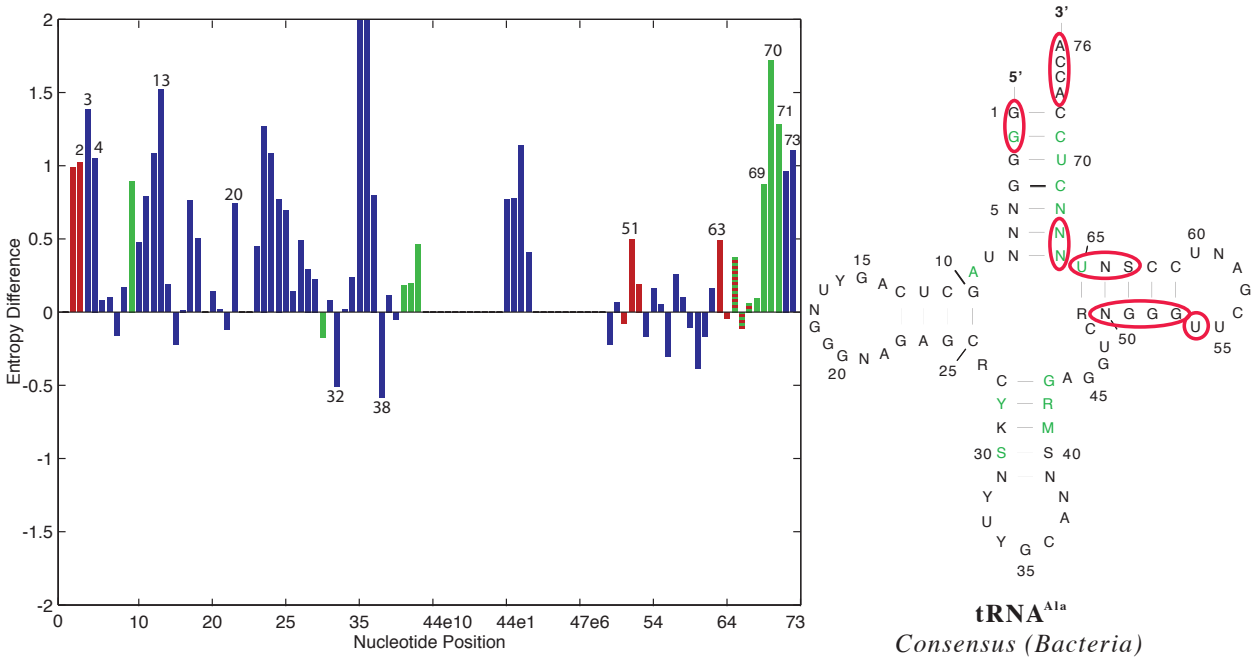


Figure 2.2: Entropy differences for bacterial tRNA^{Ala} over all isoacceptors. Green and red peaks are predicted to bind the aaRS and EF-Tu respectively. Striped peaks bind both enzymes. The nucleotides at the interface are colored accordingly on the cloverleaf schematics. The nucleotides at the AlaRS·tRNA^{Ala} interface were previously predicted by partial digestion [153]. The nucleotides at the interface with EF-Tu were derived by analogy with the EF-Tu·tRNA^{Cys} crystal structure [100]. Labeled peaks are identity/recognition elements identified by experiment or tuning elements by analogy to [154] (see Results).

A second example of the entropy difference analysis successfully pinpointing known identity elements is seen with bacterial tRNA^{Cys}. G34, C35, A36, U73 are nearly universal for tRNA^{Cys} (see Figure 2.3) and have been identified as essential for aminoacylation[155, 156]. R13·A22 in the core have also shown to have significant

information content as well known to be identity elements for bacterial $tRNA^{Cys}$ by experiment [157]. C11·G24 and G30·C40 also have significant information content, but have not been tested by experiment. G24 and C40 were recently demonstrated to be structurally and energetically important in CysRS·$tRNA^{Cys}$ binding (see Chapter 4). These nucleotides were missed previously, are recommended as subjects for future mutation studies. G15·G48 is the Levitt pair that was shown to be an identity element in $tRNA^{Cys}$ in *E. coli*, but is most commonly G15·C48. Accordingly, these positions have low information, indicating that this identity element is organism specific.

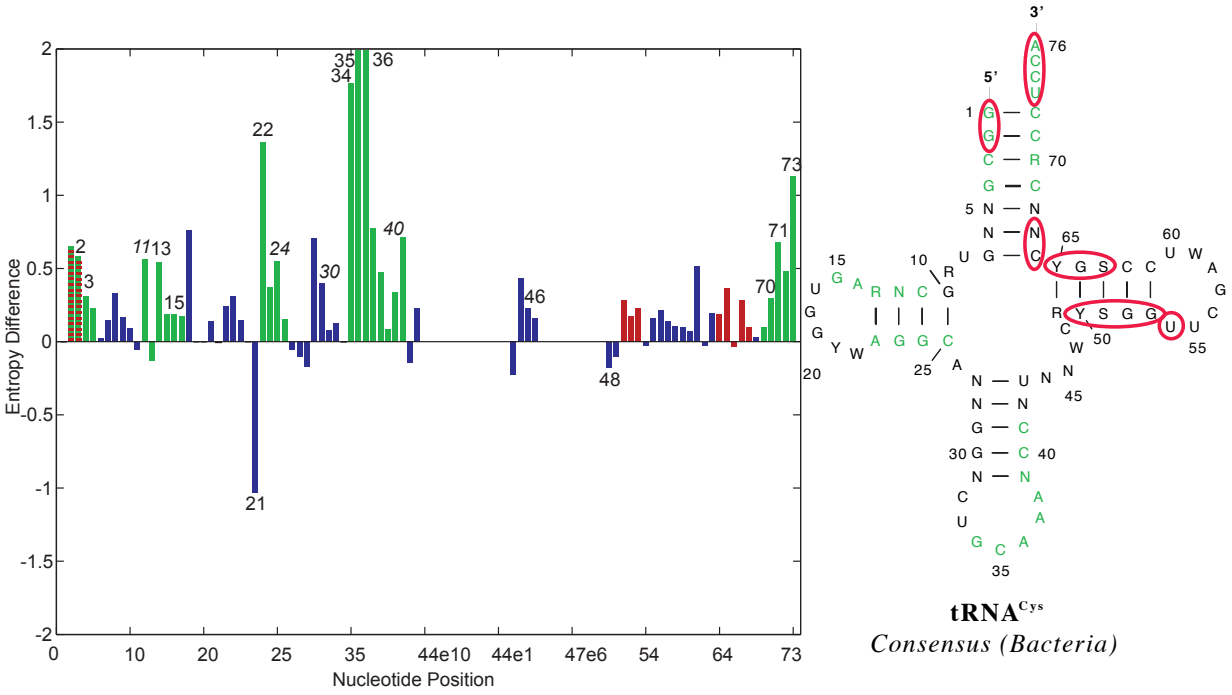


Figure 2.3: Entropy difference plot for bacterial $tRNA^{Cys}$. Green and red peaks bind the aaRS and EF-Tu respectively. Striped peaks bind both enzymes. The nucleotides at the interface are colored accordingly on the cloverleaf schematic. The nucleotides at the CysRS·$tRNA^{Cys}$ interface were predicted from the modeled *E. coli* structure (Chapter 4). The nucleotides at the interface with EF-Tu were derived from the EF-Tu·$tRNA^{Cys}$ crystal structure [100]. Labeled peaks are identity elements determined by experiment (see Results). Italic peak labels correspond to the C11·G24 and G30·C40 base pairs described in the text.

2.4.2 Prediction of identity/tuning/recognition elements

The value of the entropy difference analysis is to identify uniquely conserved nucleotides in each tRNA domain/specificity, creating a list of nucleotides for further analysis (see Tables 2.2-2.7). Some of these nucleotides are already known to be important for binding to the aaRS, EF-Tu, or to other enzymes along the translational pathway. However, many have unknown functionality. It is likely that they also perform a role in translation, whether as hitherto unknown identity/tuning elements, recognition elements for tRNA modification enzymes or the ribosome, antirecognition elements that prevent binding to the wrong aaRS, or integral parts of tertiary structure that maintain a recognizable tRNA shape for indirect readout [71]. The purpose of this list is therefore to be a reference for researchers performing mutation studies on tRNA. Mutation of all 76 nucleotides is often not practical and scientists must select a subset of nucleotides. These nucleotides are shown to be preserved through evolution and therefore would be a likely first subset of planned mutations. They are also valuable for those researchers engineering orthogonal tRNA·aaRS pairs for expanding the genetic code [158] and wish to inhibit binding of the engineered tRNA by the host aaRSs. The list was generated by identifying those nucleotide positions with an entropy difference value greater 0.75 or less than -0.375. This cutoff was selected to be the most discriminating while also highlighting most of the known identity elements in the various tRNA domain/specificities. The negative cutoff was selected to be half of the positive value due to the trend of negative peaks having smaller absolute values. One note of caution when using these tables is that as new genomic information becomes available, the sequence alignments and entropy difference analyses will need to be updated and the list may change. Another is that it is based on genomic data and therefore the C-to-U editing and other changes to the sequence made during tRNA maturation are absent from the analysis (see Chapter 1). Also, this list will not include those nucleotides that are identity/tuning/recognition elements only for a particular organism. The full set of entropy difference graphs is included in Appendix A along with the complete lists of known identity elements.

Specificity	DOL	High Information Nucleotides
Ala	Arc	G1 G2 **G3** C4 G7 U12 C13 G15 N20a G22 A23 G26 C27 C31 G35 C36 G39 G43 A44 G46 C48 G51 G52 U60 C62 C63 C66 G69 **U70** C71 C72 **A73**
Ala	Bac	**G1 G2 G3 G4** A9 C11 U12 C13 U17 G22 A23 G24 **Y32** G35 C36 A37 **N38** A44 G45 G46 U60 **C69 U70 C71 C72 A73**
Ala	Euk	G1 G2 **G3** U20 G35 C36 A37 G39 G43 A44 G45 G46 **U70** C71 C72 A73
Arg	Arc	**G1** C4 N17a **A20 C35 G36** A38 G51 C63 G69 **C72 G73**
Arg	Bac	G15 **C35 K36** Y48
Arg	Euk	C4 U16 **C35 K36** A44 G69
Asn	Arc	G1 C2 **C3** G4 A9 C11 U12 **C13** G15 N17 A20 N20a **G22** A23 G24 G31 **G34 U35 U36** A37 A38 C39 C40 G45 **G46 U47** C48 G59 G64 C69 **G70** G71 C72 **G73**
Asn	Bac	A9 U12 C13 U20 N20a G22 A23 C32 **G34 U35 U36** A37 A38 G46 **G73**
Asn	Euk	C3 U4 C11 U16 U20 U20a G24 C25 G30 R31 C32 G34 U35 U36 A37 A38 Y39 C40 G41 A44 G45 G46 U47 G70 G73
Asp	Arc	G1 C2 C3 C4 **G6** G7 U11 U13 G15 C23 A24 U25 G29 C31 C32 **G34 U35 C36** G39 C41 K43 G45 A46 C48 C50 G51 G52 G53 A59 U60 C61 C62 C63 G64 C66 **C67** G69 G70 G71 C72 **G73**
Asp	Bac	C27 **G34 U35 C36** G43 A44 G45 G49 G51 C56 C63 U65 **G73**
Asp	Euk	U5 K6 U11 U13 N17 N20a A24 C31 **G34 U35 C36 C38** G41 G45 A46 N47 C49 G51 C63 G65 R68 **G73**

Table 2.2: High information nucleotides for $tRNA^{Ala}$, $tRNA^{Arg}$, $tRNA^{Asn}$, and $tRNA^{Asp}$ grouped by domain/specificity. The nucleotides with an entropy difference above 0.75 or below -0.375 are listed along with the consensus nucleotide corresponding to that position. Nucleotides already known to be identity/tuning/recognition elements are shown in bold. See [65, 67, 119, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 5, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 5, 188] for element references.

Specificity	DOL	High Information Nucleotides
Cys	Arc	G1 C2 C3 G6 G7 N21 A22 G24 G30 C32 **G34 C35 A36** A38 C40 U60 C66 C67 G70 G71 C72 **U73**
Cys	Bac	N17 **W21 A22 G34 C35 A36** A37 **U73**
Cys	Euk	C11 **C13** U16 N17 U20 N20a **G22** G24 A31 **G34 C35 A36** 38 U39 U47 C50 W59 Y62 G64 U66 C72 **U73**
Gln	Arc	**A1** G2 C4 C5 G7 G12 U13 G15 Y17 C20a C23 N26 G29 G30 **U34 U35 G36** C40 C41 G45 A46 N47 C48 A59 U60 C66 G68 G69 C71 **U72 A73**
Gln	Bac	**N9 Y34 U35 G36**
Gln	Euk	G2 U13 N17 U20 C25 N26 A31 Y34 T35 G36 U39 A46 N47 G65 C71
Glu	Arc	C2 C4 U11 U13 C20a A20b C23 A24 U25 G29 C31 C32 **Y34 U35 C36** G39 C41 G45 A46 N47 C48 G51 A59 U60 C62 C63 **G68** G69 A70 G71 C72 A73
Glu	Bac	**C3** C5 C12 G23 **Y34 U35** C36 N38 G68 **G70**
Glu	Euk	U1 C2 C3 U11 U13 U20a A24 Y34 U35 C36 G45 G51 U60 C63 G70 G71 A72
Gly	Arc	G1 C2 G3 U11 U13 A24 N26 C31 **C35 C36** G39 C41 N47 C50 G51C63 G64 C70 G71 C72 **A73**
Gly	Bac	**G3** R15 **C35 C36** Y48 G51 **C70 G71 C72 U73**
Gly	Euk	U11 U13 N17 U20 A24 U25 N26 S30 **C35 C36** A37 S40 G45 G51 U60 C63 **G71 A73**

Table 2.3: High information nucleotides for $tRNA^{Cys}$, $tRNA^{Gln}$, $tRNA^{Glu}$, and $tRNA^{Gly}$ grouped by domain/specificity. The nucleotides with an entropy difference above 0.75 or below -0.375 are listed along with the consensus nucleotide corresponding to that position. Nucleotides already known to be identity/tuning/recognition elements are shown in bold. See [65, 67, 155, 189, 190, 191, 192, 193, 194, 195, 196, 197, 5, 198, 199, 172, 200, 201, 202, 203, 123, 100, 154, 204, 148, 205, 206, 207, 208, 209, 189, 210, 211, 212, 213, 214, 212, 215] for element references.

Specificity	DOL	High Information Nucleotides
His	Arc	C2 C3 U13 G15 **G29** G34 U35 G36 G37 **C41** G45 **C50** G51 W59 U60 C63 **G64** G65 G70 G71 C72 **C73**
His	Bac	**G-1 G34 U35 G36** Y60 **C73**
His	Euk	G1 C3 U11 U13 G15 U16 N17 U20 A24 Y27 G31 U32 **G34 U35 G36** G37 C38 C39 G40 A46 C49 N52 G57 U59 U60 G64 G65 G70 C72 **A73**
Ile	Arc	G2 G3 C5 G7 C11 U12 Y13 N20a A21 G22 A23 G24 **C27** Y28 C29 C32 **G34 A35 U36** A37 A38 G41 G42 **G43** A44 G45 G46 U47 A59 U60 C63 C66 G68 C70 C71 **A73**
Ile	Bac	A9 C11 **U12** C13 U20 G22 **A23** G24 C25 S30 **A35 U36** G46 U47 **A73**
Ile	Euk	G1 C4 C13 U20 U27 G29 N30 C32 **A35 U36** A38 C41 A44 G46 U47 G69 C72 A73
Leu	Arc	G1 C2 G3 G5 G6 G7 N9 G10 C11 C12 G15 N17a A22 G23 G24 C25 **G26** A35 R36 **S44e2 N44e3 C44e4 A46 G47 N47e4 S47e3** C48 G51 A59 U60 C63 C66 C67 C68 C70 G71 C72 **A73**
Leu	Bac	**A20a** G21 A22 A35 R36 A38 U44e1 G44e2 N44e3 N44e4 N44e5 N44e6 N44e11 N47e10 N47e9 N47e6 N47e5 N47e4 N47e3 C47e2 G47e1 **A73**
Leu	Euk	G1 G9 G10 C12 G13 N17b **G19** U20 A22 G23 G26 U33 **A35** R36 **S44e2 N44e3 N44e4 N47e4 N47e3 N47e2 G47e1** G49 G51 A59 C63 C65 C72 **A73**

Table 2.4: High information nucleotides for $tRNA^{His}$, $tRNA^{Ile}$, and $tRNA^{Leu}$ grouped by domain/specificity. The nucleotides with an entropy difference above 0.75 or below -0.375 are listed along with the consensus nucleotide corresponding to that position. Nucleotides already known to be identity/tuning elements are shown in bold.
See [65, 67, 189, 211, 216, 217, 218, 219, 220, 221, 222, 223, 224, 172, 225, 31, 226, 227] for element references.

Specificity	DOL	High Information Nucleotides
Lys	Arc	G1 G2 G3 C4 C5 A9 C11 U12 Y13 G15 A23 G24 C25 G26 G30 R31 C32 Y34 **U35 U36** A37 A38 C40 G45 G46 C48 G51 A59 U60 C63 G68 G69 C70 C71 C72 **G73**
Lys	Bac	A9 C11 U12 C13 N20a G22 A23 G24 C32 **Y34 U35 U36** A37 A38 G46 U60 **A73**
Lys	Euk	C2 C3 A9 C11 U12 C13 U16 G22 A23 G24 G30 A31 C32 **Y34 U35 U36** A37 A38 U39 C40 G46 **G51 C63** G70 G71
Met	Arc	G15 N17a G26 C28 G31 C32 **C34 A35 U36** A37 A38 C39 G42 G45 C48 A59 A73
Met	Bac	**C34 A35 U36 A73**
Met	Euk	C13 G26 **U33 C34 A35 U36 A38** C40 A44 G45 G46 U47 W54 N60 **A73**
Phe	Arc	G1 C2 C3 A9 C11 U12 **C13** N17a G20 **A22** G23 G24 C25 C27 C32 **G34 A35 A36** G37 A38 G43 G46 U47 C49 C50 G64 G65 C67 G70 G71 C72 **A73**
Phe	Bac	A9 C11 U12 C13 **U20** A23 G24 **C25 G34 A35 A36 A73**
Phe	Euk	G1 G4 A9 U12 C13 U16 U17 G22 A23 **G30 A31** C32 **G34 A35 A36 G37 U39** G46 U47 Y50 K52 U59 G63 G64 G65 C69 C72 **A73**
Pro	Arc	**G1** G2 G3 C5 C6 G7 U13 G15 N26 **G35 G36** G37 N38 G43 G45 A46 N47 C48 A59 U60 C66 G67 G68 C70 C71 **C72 A73**
Pro	Bac	G3 G7 **G15 N17a G35 G36** G46 **C48 G49** C70 **G72 A73**
Pro	Euk	G1 U11 U13 N17 U20 A24 C27 G30 C31 U32 **G35 G36** G37 U38 G39 C40 G43 G45 G46 G64 C72 **C73**

Table 2.5: High information nucleotides for $tRNA^{Lys}$, $tRNA^{Met}$, $tRNA^{Phe}$, and $tRNA^{Pro}$ grouped by domain/specificity. The nucleotides with an entropy difference above 0.75 or below -0.375 are listed along with the consensus nucleotide corresponding to that position. Nucleotides already known to be identity/tuning/recognition elements are shown in bold. See [65, 67, 228, 229, 230, 231, 232, 233, 234, 235, 5, 236, 237, 205, 175, 176, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 225, 255, 256, 257, 258, 259, 260, 161, 261, 189, 262, 263, 264, 265, 266, 267, 254] for element references.

Specificity	DOL	High Information Nucleotides
Ser	Arc	**G1 G4** 6 C12 G15 N17a G23 G24 **G26 G30** G35 A36 A38 G44e2 N44e3 N44e4 S44e5 N44e6 G47e4 S47e3 C47e2 K47e1 C48 G51 A59 U60 C63 C67 **C69 C72 G73**
Ser	Bac	**G2 S10 N20b S25** G35 A36 A37 N38 **Y44e1 R44e2 N44e3 N44e4 N44e5 N44e6 N44e7 N44e8 N44e9 N44e10 N44e11 N47e10 N47e9 N47e7 N47e6 N47e5 S47e4 N47e3 N47e2 N47e1** G51 A59 C63 **U70 C71**
Ser	Euk	**G1** G9 G10 C11 G13 N17 U20 A22 G24 C25 G26 A31 G35 A36 A38 U39 **G44e2 N44e3 S44e4 N44e5 N47e10 N47e5 S47e4 N47e3 N47e2** C48 G49 A59 **C72 G73**
Thr	Arc	**G1 C2 C3** U12 G15 N17a A23 G24 M31 **G35 U36** A38 K39 A44 G45 U47 C48 G51 C63 **G70 G71** C72 **U73**
Thr	Bac	**G1 C2** A9 U12 C13 N20a G22 A23 Y32 **G35 U36** A37 A38 G45 **G71 C72**
Thr	Euk	**G1** U20 N20a G24 C25 **G35 U36** A37 A38 G45 Y62 G70 G71
Trp	Arc	G1 G2 G3 C6 U13 N17a G29 G30 G31 C32 **C34 C35 A36** C41 G45 G46 G51 U60 C63 G67 C70 C71 C72 **A73**
Trp	Bac	**G3** A14 N20a G29 G30 U31 **C34 C35 A36** A37 A38 C40 C41 **C70 G73**
Trp	Euk	C13 N17 U20 N20a G22 U27 G30 A31 **C34 C35 A36** G45 G46 U47 G52 C62 K64 U71 **C72**

Table 2.6: High information nucleotides for $tRNA^{Ser}$, $tRNA^{Thr}$, and $tRNA^{Trp}$ grouped by domain/specificity. The nucleotides with an entropy difference above 0.75 or below -0.375 are listed along with the consensus nucleotide corresponding to that position. Nucleotides already known to be identity/tuning elements are shown in bold. See [65, 67, 228, 231, 268, 200, 269, 270, 271, 272, 273, 274, 5, 275, 276, 277, 254, 278, 230, 279, 280, 281, 172, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299] for element references.

Specificity	DOL	High Information Nucleotides
Tyr	Arc	**C1** C2 C3 G4 C6 A9 U12 C13 G22 A23 S29 G30 A31 C32 **G34 U35 A36** G37 A38 U39 C40 S41 A44 G46 U47 C50 A59 U60 C63 G64 G67 C69 G70 G71 **G72 A73**
Tyr	Bac	**G1** G2 N9 S10 A14 Y20a S25 C28 G30 C32 **G34 U35** A36 C40 G42 **G44e2 N44e3 N44e4 N47e6 N47e5 N47e4 C47e3 U47e1** G49 C71 **C72 A73**
Tyr	Euk	**C1** C5 U16 U17 N20a C25 G30 A31 C32 **G34 U35** A36 A38 U39 C40 G45 G46 U47 C50 G64 G68 **G72 A73**
Val	Arc	G1 G2 G3 C4 G7 U11 C12 U13 G15 U20a G23 A24 C28 C31 **A35 C36** G39 G42 R43 G45 U47 S52 A59 C66 G68 G69 C70 C71 C72 **A73**
Val	Bac	A9 U12 C13 G22 A23 **A35 C36** N38 G46 U47 U60 **A73**
Val	Euk	U11 U13 U20 A24 U27 C31 **A35** C36 C38 A44 G45 G46 U47 C48 G64 **A73**

Table 2.7: High information nucleotides for $tRNA^{Tyr}$ and $tRNA^{Val}$ grouped by domain/specificity. The nucleotides with an entropy difference above 0.75 or below -0.375 are listed along with the consensus nucleotide corresponding to that position. Nucleotides already known to be identity/tuning/recognition elements are shown in bold.
See [65, 67, 161, 300, 268, 301, 302, 249, 303, 304, 305, 306, 307, 308, 60, 172, 309, 225, 310, 311, 312, 313, 5, 314] for element references.

2.4.3 Confirmation and prediction of tuning elements

Tuning elements are tRNA nucleotides that assist in binding to EF-Tu. TRNAs differ in their respective binding energies when they each have the same amino acid, indicating that the tRNA body and amino acid set are tuned so that the binding energy is well-balanced for docking/undocking to the EF-Tu [101, 102]. These nucleotides may be easier to identify because most variations will be on the side of the tRNA (for aaRS·tRNA docking, the aaRS can contain coevolving residues to bind identity elements). A single EF-Tu must

dock all tRNA species in an organism, and at least 50% of the protein·RNA interface residues on EF-Tu are conserved across all domains of life and 80% within *Bacteria* [63]. Also, the tRNA sequences may differ between organisms or domains of life, but each will bear the cognate amino acid and it is this ligand that accounts for a significant portion of the total binding energy (between 20% and 40%) [102, 63]. The nucleotides that interact with EF-Tu in the available structures [99, 100, 154] are at positions 1, 2, 50-54, 63-67, and 73. The G51·C63 base pair has the strongest signal in the entropy difference analysis and therefore has been shown on each figure in Appendix A. Other G·C base pairs in these regions have been identified as tuning elements. Due to structural concentration and similarity of tuning elements, the entropy difference analysis is not sufficiently sensitive to locate tuning elements other than the G51·C63 base amidst the strongly conserved identity elements. Consensus sequence comparison is therefore used in the following sections to predict the location of tuning elements among the different tRNA domain/specificities.

Chemically similar amino acids often have the same discriminator base, indicating that this base may allow detuning between similar amino acids charged on the same tRNA species. The discriminator base shows the highest correlation to amino acid specificity and is conserved over all three domains of life in eleven specificities: A for $tRNA^{Ala}$, $tRNA^{Ile}$, $tRNA^{Leu}$, $tRNA^{Met}$, $tRNA^{Phe}$, $tRNA^{Tyr}$, and $tRNA^{Val}$ (mostly hydrophobic); U for $tRNA^{Cys}$; G for $tRNA^{Asn}$, $tRNA^{Asp}$, and $tRNA^{Ser}$. $tRNA^{His}$ (archaea/bacteria) is the only tRNA with a C for a discriminator base.

Base pairs on the acceptor stem that interact with EF-Tu also show high correlation to amino acid specificity. The base pair G1·C72 is conserved in twelve specificities and several other domain/specificities. Those with the A1·U72 base pair often use it as an identity element. Base pair 2·71 is conserved in seven specificities: G·C or C·G in $tRNA^{Leu}$ (archaea/bacteria), and $tRNA^{Met}$; G·C in $tRNA^{Ala}$ (see Figure 2.2), $tRNA^{Gln}$, and $tRNA^{Pro}$; and C·G in $tRNA^{Gly}$, and $tRNA^{Thr}$. Base pair 3·70 is also conserved in seven specificities: C·G or U·A in $tRNA^{Asn}$, $tRNA^{Asp}$, and $tRNA^{Thr}$; G·C or A·U in $tRNA^{Gly}$, $tRNA^{Pro}$ (archaea/bacteria), and $tRNA^{Trp}$; and G·U in $tRNA^{Ala}$. Finally, the common arm

```
Ala Ar GGGCYSGUAGNUCAGCNN-GGU---AGANYGCCNCCUUKGCAAGGNGG-----------AGGC----------CNCGGGUUCAAAUCCCGNCSRGUCCA
Ala Ba GGGGNNNUAGCUCAGYU--GGG---AGAGCRCYKSNYUYGCANNSMRG-----------AGGU----------CRNGGGUUCGANUCCSNUNNNCUCCA
Ala Eu GGGSNUGURGYKYAGU---GGU---AKMRCGCUYGCUUYGCAUGCRAG-----------AGGU----------CCNGGGUUCGAUUCCCNGCNNSUCCA
Arg Ar GGGCNSGURGSNUAGYNN-GGAU--AKNSYRNNGGSCUNCGRANCCNN-----------WGRN----------CSCGGGUUCRAAUCCCGSCSGGCCCG
Arg Ba GNNYYNGUAGCUCARYN--GGAU--AGAGCRNNNGNYUNCKAANCNNN-----------NGGU----------YGNRGGUUCGANUCCYNCCGGGNNCR
Arg Eu GNYCNNGURGCNYAAU---GGWN--ARNGCRUYNGNCUNCKRANCNRA-----------AGRU----------UGYGGGUUCGANUCCYGCCNNGGKYK
Asn Ar GCCGCNGUAGCUCAGYN--GGU---AGAGCGNCNGGCUGUUAACCNGN-----------NGGU----------CSCMGGUUCRAGUCCKGNCNGCGGCG
Asn Ba UCCNNNNUAGCUCAGUN--GGU---AGAGCRNNNGRCUGUUAAYCNNN-----------NGGU----------CGYNGGUUCGAGUCCNRCNNNNGGAG
Asn Eu GYCUCNGURGCKCAGUN--GGUU--AGMGCGUNCGRCUGUUAAYCGNA-----------AGGU----------CGNKRGUUCGAKYCYMNCYNGRGRCG
Asp Ar GCCCNGGURGUGUAGN---GGCCU-AUCAUGNRGSCCUGUCRMGSCYK-----------YGA----------CYCGGGUUCRAAUCCCGGCCNGGGCG
Asp Ba GGNSYNGURGYKYAGUN--GGUU--ARMRYGCCNSCCUGUCACGSNGG-----------AGRU----------CGCGGGUUCGANUCCCGUCNGNNCCG
Asp Eu UCCNUKNUAGUWUARY---GGUN--AGWAUWCNCGCUUGUCACGCGGG-----------AGA----------YCSGGGUUCGAUUCCCSGNMRSGGAG
Cys Ar GCCRRGGURGCNGAGY---GGC---NANGCGGCGGMCUGCAGAKCCGN-----------WNNA----------CSCSGGUUCAAAUCCSGSCCYYGGCU
Cys Ba GGCGNNGURGCNRAGU---GGY---WAGGCANNGGNCUGCAAANCCNU-----------NNW----------CRYSGGUUCGAWUCCSGYCNNCRCCU
Cys Eu GGGNNNAUAGCUCAGU---GGU---AGAGCANNNGAUUGCARAUCNNU-----------AGGU----------CNCNRGUUCGAWYCYGGNUGNSCCCU
Gln Ar AGYCCNGURGUGUAGY---GGYCA-AUCAUNCNGGNCUUUGGANCCNG-----------NGA----------CSSCGGUUCGAAUCCGSSCNGGRCUA
Gln Ba UGSSSNRUNGYSWARY---GGY---ARSRCANCNGNYUYUGRNYCCGK-----------NRNN----------YSNAGGUUCGAAUCCUNSYNSSSCAG
Gln Eu GGUUNYRURGUGUARU---GGUU--AKCACNYYRGAYUYUGAWUCCRR-----------NAA----------YCYGRGUUCGAAUCYCGGYRNGACCW
Glu Ar GCUCCNGURGUGUAGYCC-GGCCA-AUCAUGCNGGCCUYUCRMGSCNG-----------NGA----------CYSGGGUUCAAAUCCCGGCNGGAGCA
Glu Ba GNCCCNUUSGUCWAGY---GGYY--ARGACRYCNSNYUYUCANGSNGG-----------NAA----------CRSGGGUUCGAWUCCCSUWSGGGNYA
Glu Eu UCCGWURURGUSUARY---GGYU--AGSAYWYNUSGCUYUCACSSAGR-----------AGR----------CCSGGGUUCGAUUCCCSGYAWSGGAR
Gly Ar GCGSCNRURGUSUAGYNN-GGU---AKSACNNGGGCYUNCCAAGSCNN-----------URA----------CCCGGGUUCRAAUCCCGGYNGNCGCA
Gly Ba GCGGNNRUAGYUYARU---GGU---ARARCNYCAGCYUKCCAAGCUGR-----------AGR----------YGCGRGUUCGANUCCCGUYNNCCGCU
Gly Eu GCRYNNKUGGUNUAGU---GGUN--ARNAUNNNWSCYUNCCAWGSWNK-----------NGA----------CCCGGGUUCGAUUCCCGKNNNNYGCA
His Ar GCCGGGGURGKGUAGY---GGUU--AKCMYGNNGGACUGUGGAKCCNN-----------NGA----------CCCGGGUUCRAWUCCCGGCCCCGGCC
His Ba GYGGNUGUAGCUCAGUN--GGUN--ARAGCRCYRGWUUGUGGNNCYGG-----------NGGU----------CGNGGGUUCRANYCCCNUCRNYCRCC
His Eu GCCGNNAUAGUNUAGU---GGUN--AGNACNYNACGUUGUGGCCGUNG-----------MRA----------CCCNGGUUCGAUUCCNGGUNNCGGCA
Ile Ar GGGCCMGUAGCUYAGCYW-GGUN--AGAGCGCYCGGCUGAUAACCGGG-----------AGGU----------CMNGGGUUCGAAUCCCNKCKGGCCCA
Ile Ba GGGCNURUAGCUCAGUU--GGUU--AGAGCRCNCSNCUGAUAANSGNG-----------AGGU----------CSNNGGUUCRANUCCNNNYRGGCCCA
Ile Eu GSUCYNKUAGCUCAGUN--GGUU--AGAGCGUSGNKCUWAUAASGCGA-----------AGGU----------CGNGGGUUCGANCCCCNCNKGGASCA
Leu Ar GCGRGGGUNGCCRAGCNN-GGYCA-AAGGCGNMRGRCUYARGAYCYKNUSNC-------GUAG-----GNSUUCGNGGGUUCRAAUCCCNYCCCYCGCA
Leu Ba GCSSNNGUGGCGRAAYU--GGUA--GACGCRCNNGNYUNARRANCNNGUGNNN------KNNN-----NNNCGUGNGGGUUCRANUCCCNYCNNSSGCA
Leu Eu GGNANKAUGGCCGAGU---GGUYN-AAGGCGNYRGRYUYARGUYCURNUSNN-------NNWN------NNNGCGYGGGUUCGAAUCCCRCUNNUGUCA
Lys Ar GGGCCNGUAGCUYAGUCA-GGY---AGAGCGNCKGRCUYUUAAYCMGN-----------NGGU----------CGNGGGUUCAAAUCCCKNCNGGCCCG
Lys Ba GSGYNNNUAGCUCAGYN--GGU---AGAGCANCNGACUYUUAAUCNGN-----------NGGU----------CSNRGGUUCGARUCCYNSNNNGCNCA
Lys Eu GCCYKGNUAGCUCARUY--GGU---AGAGCGUYNGACUYUUAAUCKRA-----------WGGU----------YGYGGGUUCGAGUCCCGCNNWGGGCK
Met Ar RSCGSGGURGSKYAGCYN-GGUN--AKMSCGCCGGGCUCAUAACCCGG-----------AGGU----------CNNGGGUUCRAAUCCCNNCCSCGSYA
Met Ba GGCNNNGUAGCUCAGYN--GGUN--AGAGCRNCNGRCUCAUAAYCNGN-----------NGGU----------CGNNGGUUCRARUCCNNCCNNNGCYA
Met Eu RGCNNNNURGCKCAGUN--GGW---AGMGCGYNGGKCUCAUAAYCCNR-----------AGGU----------CNNNRGWUCGAANCYNNSNNNNGCYA
Phe Ar GCCGNGGUAGCUCAGCYN-GGG---AGAGCGCYSGRCUGAAGAYCSGG-----------WKGU----------CCCSGGUUCAAAUCCSGGCCGCGGCA
Phe Ba GSCNNKGUAGCUCAGUN--GGU---AGAGCRNNNGACUGAARAUCNNN-----------GKGU----------CGNNGGUUCGAUUCCGNYCNNNGSCA
Phe Eu GCNGNNAUAGCUCAGUU--GGG---AGAGCGYNAGACUGAAGAUCNKR-----------AGGU----------CCYSKGUUCGAUCCMGGGUNNCNGCA
Pro Ar GGGRCCGURGKSUAGCYU-GGUN--AUSMUNCNGGSUUNGGGNSCCNG-----------WGAN----------CCCSGGUUCRAAUCCSGGCGGYCCCA
Pro Ba CGGGRYGUAGCGCAGYNN-GGU---AGCGCRCYNGNYUNGGGANCNRG-----------RGGU----------CGNRGGUUCRAAUCCYGYCRYCCCGA
Pro Eu GGCYNNNUGGUCUAGU---GGU---AUGAUNCYYGCUUWGGGUGCRRG-----------WGGU----------CCCGGGUUCRAWUCCCGGNNNAGCCC
Ser Ar GCCGRGRURGCCNAGCNN-GGUN--AAGGCGCNGGNCUGGAAANCCNGUGNNSN-----KNWN-----NGSCKCSNGGGUUCAAAUCCCNSYCYCGGCG
Ser Ba GGAGRGNUGSCCGAGU---GGYYN-AAGGSRCCNGNCUGGAAANCNGKYRNNNNN----NNAN----NNSNNNCGNGGGUUCGAAUCCCNCNCYCUCCG
Ser Eu GNCRNNRUGGCCGAGU---GGUU--AAGGCGWNNGACUNGAAAUCNNWUGNSN------UUU------NSNNGCGCRGGUUCGAAUCCYGCYGNYGUCG
Thr Ar GCCUNGGURGCUCAGCYN-GGU---AGAGCGNNUSMYUNGUAAKCANN-----------AGGU----------CGCGGGUUCRAAUCCCGCCCNNGGCU
Thr Ba GCCGNYNUAGCUCAGUN--GGU---AGAGCANNNSMYUKGUAAKSNNN-----------AGGU----------CRNSGGUUCRANUCCNNYNRNNGGCW
Thr Eu GCCNNYNUAGCUYAGU---GGU---ARAGCGYYNSWCUNGUAAWSNRR-----------AGRU----------CNNGRGUUCRAUUCYSNNWGKNGGCW
Trp Ar GGGGNCRUGGCGUAGYCN-GGU---AKCGYGGCGGGCUCCAGAMCCSN-----------YGGW----------CSKGGGUUCAAAUCCCASYGNCCCCA
Trp Ba AGGGGNNUAGYUYAAYN--GGU---AGARCRNCGGUCUCCAAAACCGN-----------NGGN----------UGNGGGUUCGANUCCUNCNNCCCCUG
Trp Eu GRSNNCKURGCKCARU---GGU---AGMGYRUNNGAYUCCARAUCNNA-----------AGGU----------YGNRGGUUCAANUCCYKUMGNNSUCA
Tyr Ar CCCGCCKUAGCUCAGYYN-GGY---AGAGCGSCSGACUGUAGAUCSGS-----------AKGU----------CSCNGGUUCAAAUCCGGSMGGCGGGA
Tyr Ba GGRGRGGUNSCNRAGU---GGYYA-AASGSRGCNGACUGUAAAUCNGYYGNN-------KNNN-----NNCUUCGNAGGUUCGAAUCCUNCCCYCYCCA
Tyr Eu CCSNCNWUAGCUCAGUU--GGU---AKAGCRUNNGACUGUARAUCNNW-----------WGGU----------CGCUGGUUCGAWUCCGGCWSGNSGGA
Val Ar GGGCYCGUGGUCUAGNN--GGYU--AUGAYGYCGCCYUKACAMGGCGR-----------AGRU----------CNSSGGUUCRAAUCCSNNCGGGCCCA
Val Ba GGGCSNNUAGCUCAGYN--GGK---AGAGCRCYNSCYUKACANGGNRG-----------RGGU----------CRNNGGUUCGANUCCNNYNNNGCCCA
Val Eu GGUYNNRURGUSUAGU---GGUU--AUSACGUCUSCYUNACACGSAGA-----------AGGU----------CSCNRGUUCGAWCCYSGSYNGGANCA
```

Figure 2.4: Consensus sequences of all tRNA domain/specificity groups. Dashed positions are less than 20% populated. The expanded nucleotide code is: R (purine A/G), Y (pyrimidine C/U), S (strong G/C), W (weak A/U), K (ketone G/U) and M (amino A/C).

base pair G51·C63 is highly conserved across all domains of life in $tRNA^{Glu}$, $tRNA^{Asp}$, $tRNA^{Gly}$,$tRNA^{Leu}$, and $tRNA^{Ser}$ (archaea/bacteria), and shows higher information content than the average. Among the other specificites ranging from strong binding to weak binding tRNAs, the archaeal versions are the most likely to contain the G51·C63 base pair – $tRNA^{His}$ (archaea/bacteria), $tRNA^{Thr}$ (archaea), $tRNA^{Ala}$ (archaea), $tRNA^{Cys}$ (archaea), $tRNA^{Lys}$ (archaea/eukarya), $tRNA^{Arg}$ (archaea/eukarya), and $tRNA^{Trp}$ (archaea/bacteria) – which could be indicative of an archaeal specific feature of the EF-Tu. In general, the stronger binding tRNAs [101] have a larger number of conserved bases at the binding interface, indicating that these base positions could be tuning the binding affinity of the aa-tRNA to EF-Tu.

2.5 Conclusion

Bioinformatics methods reveal that uniquely conserved nucleotides are charcteristic within tRNA domain/specificity and can be predicted using Shannon entropy differences and consensus sequence comparisons. Many high information nucleotides are consistent with previous experiments, as shown with the well studied bacterial $tRNA^{Ala}$, indicating that this measure can successfully identify nucleotides that have been conserved to bind a specific enzyme. The bacterial $tRNA^{Cys}$ provides a different example with several high information nucleotides validated as known identity elements while other high information nucleotides, specifically C11·G24 and G30·C40, are predicted to be additional identity elements. For many tRNA domain/specificities, particularly in the *Archaea*, little or no information on identity/tuning/recognition elements is available and the lists of high information nucleotides can serve as a basis for designing future studies. These lists can also facilitate the identification of recognition elements at the ribosome or to enzymes responsible for tRNA maturation, areas that are only sparely explored. Tuning elements are more readily identifiable through consensus sequences of the well-balanced evolutionary profiles. Here, the discriminator base is shown to be conserved in 11 of the 20 specificities

and the characteristic G51·C63 base pair is conserved in five specificities over all three domains of life. Archaeal organisms have a greater propensity to contain this base pair than tRNAs of the same specificity in other domains of life, the significance of which is an area for future study.

2.6 Acknowledgements

The author thanks Dan Wright, John Eargle, Elijah Roberts, and Anurag Sethi for many helpful discussions and technical assistance. Funding was provided by: NSF (MCB04-46227, FIBR-SBCSF0526747), and ABP was supported by NIH Chemical Biology Training Grant (5T32GM070421).

CHAPTER 3

EXIT STRATEGIES FOR CHARGED TRNA FROM GLUTAMYL-TRNA SYNTHETASES

3.1 Summary

For several Class I aminoacyl-tRNA synthetases (aaRSs), the rate determining step in aminoacylation is the dissociation of the charged tRNA from the enzyme. In this chapter, the effects of protonation state changes in the active site and the presence/absence of AMP and EF-Tu on the release of charged tRNA from the aaRS is computationally explored. Through molecular modeling, internal pKa calculations, and molecular dynamics simulations, distinct, mechanistically relevant post-transfer states with the charged tRNA (Glu-tRNA$^{\mathrm{Glu}}$) bound to glutamyl-tRNA synthetase from *Thermus thermophilus* are considered. The behavior of these non-equilibrium states is characterized as a function of time using dynamical network analysis, local energetics, and changes in free energies to estimate transitions that occur during the release of the tRNA. The hundreds of nanoseconds of simulation time reveal system characteristics that are consistent with recent experimental studies. The energetic and network results support the previously proposed mechanism in which the transfer of the amino acid to the tRNA is accompanied by the protonation of AMP to H-AMP. Subsequent migration of the proton to water reduces the stability of the complex and loosens the interface both in the presence and absence of AMP. Subsequent undocking of AMP or tRNA then proceeds along thermodynamically competitive pathways. Release of the tRNA acceptor stem is further accelerated by the deprotonation of the α–ammonium group on the charging amino acid. The proposed general base is Glu41, a residue binding the α–ammonium group that is conserved in both structure and sequence across nearly all Class I aaRSs. This universal handle is predicted

through pKa calculations to be part of a proton relay system for destabilizing the bound charging amino acid following aminoacylation. Addition of EF-Tu to the aaRS·tRNA complex stimulates the dissociation of the tRNA core and acceptor stem [1].

3.2 Introduction

The aminoacyl-tRNA synthetases (aaRSs) help maintain the genetic code by recognizing their cognate tRNAs and amino acids from the pool of competing reactants within the cell [68, 64]. In the majority of cases, the formation of the aminoacylated tRNAs (charged tRNA) within the active site occurs via a two step process (see Figure 3.1). In the first step, the amino acid is activated by ATP, forming an aminoacyl-adenylate and pyrophosphate. In the second step, the amino acid moiety on the adenylate is transferred to the $2'$ hydroxyl group at the $3'$ end of the tRNA with simultaneous formation of AMP product. AaRS·tRNA complexes before and after amino acid transfer are referred to as pre- and post-transfer states, respectively. Previous biochemical studies and crystal structures have provided valuable information about the first step of aminoacylation, the binding site of the adenylate, and the mode of interactions between identity elements on the tRNA and highly conserved residues on the aaRS (see [5] and references therein). More elusive are the details of how charged tRNA dissociates from the aaRS prior to binding elongation factor Tu (EF-Tu) for delivery to the ribosome. In Class I aaRSs, tRNA dissociation is the rate determining step for tRNA aminoacylation, which has been shown to be stimulated in the presence of EF-Tu [76, 78]. Dissociation has been hypothesized to begin with the charged $3'$ end of the tRNA exiting the active site while the anticodon remains strongly bound to the aaRS [315],

[1]The contents of this chapter are largely based on work previous published as Alexis Black Pyrkosz, John Eargle, Anurag Sethi, and Zaida Luthey-Schulten. "Exit strategies for charged tRNA from GluRS," *J Mol. Biol.*, 397, 1350-1371 © Elsevier 2010. Design of the study, joint system setup, joint parameterization, joint running of simulations, rmsd analysis, Mg^{2+} analysis, principal component analysis, joint network analysis, local energetics analysis, bioinformatics, and free energy analysis were performed by ABP. Joint system setup, joint running of simulations, joint pKa calculations, and simulation/network analysis of the EF-Tu·$tRNA^{Glu}$·GluRS·GTP complex were conducted by John Eargle. Joint parameterization, joint network analysis and joint pKa calculations were performed by Dr. Anurag Sethi.

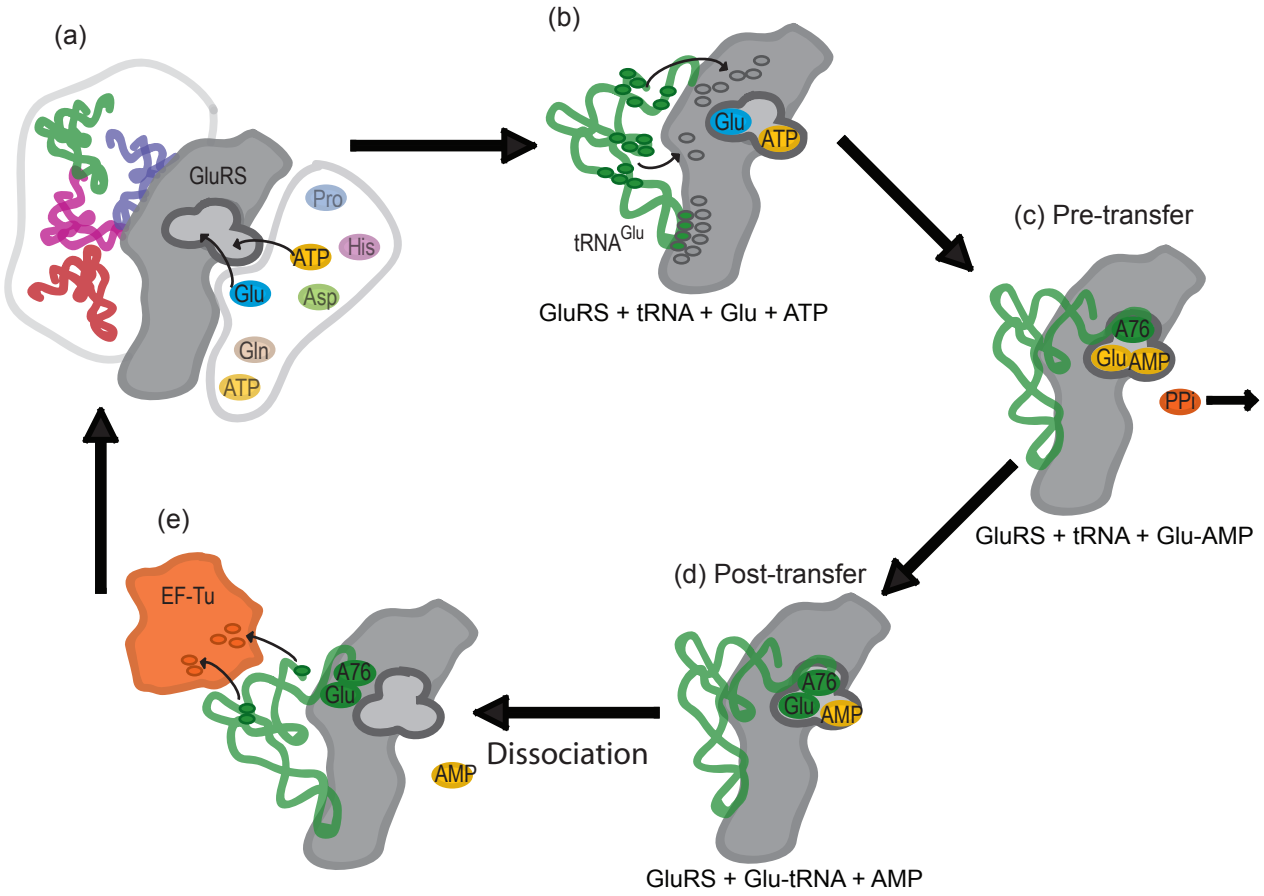


Figure 3.1: Summary of aminoacylation reaction for GluRS·tRNA^{Glu}. Panel (a) shows the apo-GluRS attracting ATP and selecting the cognate amino acid glutamate and tRNA^{Glu} from the cellular pool. Panel (b) indicates the recognition of tRNA^{Glu} identity elements (green ovals) by highly conserved residues (gray ovals) in GluRS. Panel (c) shows the Pre-transfer state(s) with formation of the adenylate. Panel (d) contains the Post-transfer state(s) with the newly charged tRNA. Panel (e) shows the charged tRNA dissociating from the synthetase before association with the EF-Tu through the tuning elements (green ovals to orange ovals) and subsequent transportation to the ribosome. The system can sample a large ensemble of states at each stage of the reaction.

In this study, the series of events occurring in the active site that control tRNA dissociation is investigated. The structure of the glutamyl-tRNA synthetase (GluRS) complexed with tRNA^{Glu} from *Thermus thermophilus* [316] was used as a representative of monomeric Class I aaRSs. Although GluRS is atypical of Class I aaRSs in that it requires tRNA to be bound before the aminoacyl adenylate can be formed, the final process of AMP

and aa-tRNA dissociation involves an analogous set of molecules in all Class I aaRSs. The modeled post-transfer states are differentiated by protonation of AMP and its neighboring amino acid residues and by the presence/absence of AMP in the active site. These states have been selected based on suggested reaction mechanisms [317] and internal pKa calculations. Through comparative analyses of each state's behavior with the pre-transfer state and experimental results, the undocking of AMP and changes in protonation states are evaluated as possible exit strategies for tRNA dissociation. The results indicate that both factors assist in the release of the charged tRNA from the enzyme.

Structure and Phylogenetic Background of the GluRS·tRNA^{Glu} Complex

AaRSs are divided into two classes based on the structurally distinct, conserved core or catalytic domain (CD) containing the active site [318, 319]. The CD of the Class I aaRSs forms a Rossman fold with a three layered $\alpha\beta\alpha$ topology containing a parallel β-sheet architecture. The active site is located at the C-terminal loops of the β-strands (see Figure 3.2a). Within the active site are the highly conserved HIGH and KMSK sequence motifs, which bind ATP during adenylate formation. The histidines in the HIGH motif form contacts with the phosphates while the KMSK loop is located near the adenine base. Located between the two halves of the Rossman fold (RF-N and RF-C), the connective polypeptide (CP1) insertion binds the 3′ end of the tRNA during aminoacylation. Class I aaRSs are further differentiated by the fold of the anticodon binding (ACB) domain [320, 321]. GluRS is in the Class Ib subgroup with a set of α-helices (four-helix junction or 4HJ) connecting the CD to the C-terminal, α-helical ACB domain. GluRS in *T. thermophilus* has been crystallized in a variety of states prior to the second step of aminoacylation [322, 323, 324, 316]. The crystal structure used for the current study contains GluRS (468 residues) with a Glu-AMP analog and transcribed tRNA^{Glu} in the active site [PDB code: 1N78] [316] (see Figure 3.2a for the structure and b for the standard tRNA cloverleaf schematic). Use of the analog creates an unreactive substrate complex mimicking the pre-transfer system state, which serves as a starting point for this study.

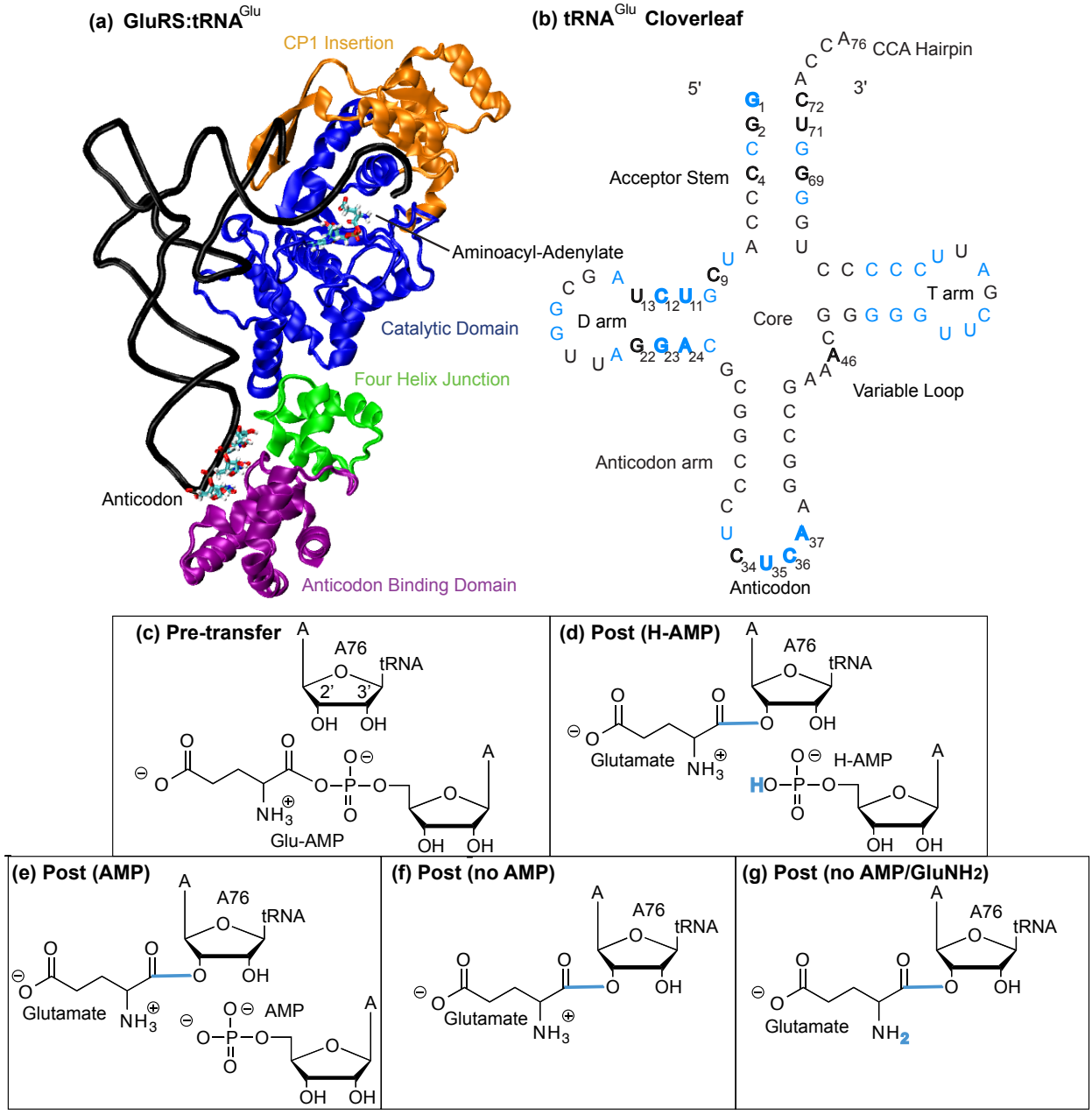


Figure 3.2: GluRS·tRNA^{Glu} system used in simulations: (a) GluRS with functional domains labeled, (b) tRNA^{Glu} cloverleaf schematic where nucleotides with greater than 75% sequence identity across the evolutionary profile are colored blue. Identity elements are in bold. Panels (c-g) show the small molecule substrate/products in the active site for the different system states. Using the Pre-transfer state as the control, the bonds and atoms colored blue in each post-transfer state indicate the changes made to the AMP or charging amino acid moiety.

GluRS has a divergent evolutionary history that has led to several classes of GluRS as well as GlnRS [325, 64, 326]. GluRS enzymes are divided into α (bacterial) and β (archaeal/eukaryal) types based primarily on the fold of their nonhomologous ACB domain structures. Both classes have discriminating (D-GluRS) and nondiscriminating versions based upon the ability of GluRS to charge $tRNA^{Glu}$ and/or $tRNA^{Gln}$ with glutamate. The α-type GluRS in some bacterial organisms has evolved specific residues to recognize the third anticodon base, discriminating between $tRNA^{Glu}$ and $tRNA^{Gln}$ [327]. Those bacteria with D-GluRS have either acquired the eukaryal-type GlnRS through horizontal gene transfer [328] or evolved a GluRS2, which specifically recognizes and misacylates $tRNA^{Gln}$ with glutamate for subsequent reduction [329, 330, 148]. Due to the lack of crystal structures containing GluRS·$tRNA^{Glu}$ in these various subgroups, the scope of the current study is limited to the discriminating α-type represented by GluRS in *T. thermophilus*.

MD Simulations and Analysis Methods for aaRSs Molecular dynamics (MD) is a powerful method that has been used to study correlations [331], signaling pathways [332, 131], editing [333], and binding free energies [334, 335, 336, 337, 338] in aaRSs. Here, long-timescale simulations of the entire GluRS·$tRNA^{Glu}$ complex with explicit solvent and both mono- and divalent ions are used to determine the dynamical and energetic behavior for the various pre- and post-transfer states. The behavior of these states is characterized as a function of time using dynamical networks, local energetics, and changes in free energy to estimate the transitions that occur during tRNA dissociation.

3.3 Methods

3.3.1 Bioinformatics

Evolutionary analyses of the structures and sequences of Class I aaRS catalytic domains have already been conducted [70, 139] and are in good agreement with phylogenetic analyses of the complete sequences [64]. The evolutionary analyses in this study were

performed to measure conservation of the residues and nucleotides important for binding either within the active site or along the GluRS·$tRNA^{Glu}$ interface. Because the ACB domain is nonhomologous between the bacterial version of GluRS and the archaeal and eukaryotic versions, the evolutionary profile was exclusively limited to bacterial sequences. Further, because the bacterial GluRS has diverged into discriminating and nondiscriminating versions, the set of sequences used to build the evolutionary profile was filtered to include only discriminating GluRS sequences containing the characteristic residue Arg358 (*T. thermophilus* numbering) [327].

D-GluRS The GluRS sequence from *T. thermophilus* was used to perform a BLAST search [339] over the NCBI non-redundant database [340] with an E-value cutoff of 10^{-5}. Alignment of all bacterial GluRS sequences was performed with ClustalW [149] and manually improved (particularly in the highly divergent 4HJ). Only sequences with arginine or similar lysine in corresponding positions to characteristic residue Arg358 in D-GluRS from *T. thermophilus* were retained. The evolutionary profile was prepared by applying SeqQR [139] with a 50% sequence identity cutoff, yielding a set of 23 sequences which are listed in Tables 3.1, representing 12 of the major bacterial phyla.

The SeqQR algorithm is used to determine the set of the most linearly independent sequences to form a statistically well-balanced profile that represents the maximum phylogenetic diversity while minimizing the number of sequences.

tRNA The sequences for bacterial $tRNA^{Glu}$ (which have CUC or UUC anticodons) were downloaded from the tRNA Compilation 2000 [341] and the Integrated Microbial Genome database at the Joint Genome Institute [145]. TRNA sequence alignment was performed using ClustalW with manual improvement. SeqQR was used with a sequence identity cutoff of 85%, yielding 56 sequences in the final evolutionary profile. The MultiSeq plugin [150] in VMD [151] was used to construct and analyze the evolutionary profiles. The organisms represented in the evolutionary profiles are provided in Tables 3.1.

Species	D-GluRS	tRNA CUC	tRNA UUC
Aquifex aeolicus			1
Acaryochloris marina			1
Acidobacteria bacterium	1		
Aeromonas hydrophila			1
Anaeromyxobacter sp.			1
Arcobacter butzleri			1
Arthrobacter aurescens	1		
Atopobium rimae	1		
Bacillus thuringiensis			1
Bacteroides uniformis		1	
Bacillus subtilis			1
Blastopirellula marina	1		
Brucella melitensis			1
Campylobacter curvus		1	
Campylobacter fetus		1	
Candidatus Carsonella ruddii			1
Candidatus Liberibacter	1		
Candidatus Pelagibacter ubique			1
Candidatus Ruthia magnifica			1
Candidatus Sulcia muelleri			1
Candidatus Vesicomyosocius okutanii			1
Chlamydia muridarum	1		
Clostridium hiranonis	1		
Clostridium sp.		1	
Clostridium tetani	1	1	
Cyanothece sp.			1
Deinococcus radiodurans	1		
Desulfatibacillum alkenivorans			1
Ehrlichia canis			1
Escherichia coli			1
Eubacterium dolichum			1
Flavobacteriales bacterium			1
Gloeobacter violaceus	1		
Hahella chejuensis			1
Heliobacterium modesticaldum	1		
Herpetosiphon aurantiacus	1		
Hyphomonas neptunium	1		
Kordia algicida		1	
Lactobacillus brevis		1	
Lactobacillus delbrueckii		1	

Continued on next page

Table 3.1: List of organisms in the evolutionary profiles for D-GluRS and $tRNA^{Glu}$.

Species	D-GluRS	tRNA CUC	tRNA UUC
Lactobacillus johnsonii			2
Legionella pneumophila	1		
Leuconostoc mesenteroides		1	
Methylibium petroleiphilum			1
Microscilla marina	1		2
Mycoplasma mycoides			1
Mycoplasma penetrans			1
Mycoplasma pulmonis			1
Nostoc punctiforme			1
Oceanospirillum sp.	1		
Opitutaceae bacterium		1	
Petrotoga mobilis		1	1
Planctomyces maris			1
Plesiocystis pacifica		2	
Propionibacterium acnes	1	1	
Prosthecochloris vibrioformis	1		
Pseudoalteromonas haloplanktis			1
Psychroflexus torquis	1	1	
Rhizobium leguminosarum		1	
Rhodopirellula baltica			1
Ruminococcus gnavus			1
Ruminococcus obeum		1	
Sphingomonas sp.	1		
Stigmatella aurantiaca	1		
Streptococcus pneumoniae			1
Thermotoga maritima	1		
Thermus thermophilus	1	1	
Thiomicrospira denitrificans			1
Treponema pallidum	1		
Victivallis vadensis			1
Wolinella succinogenes			1

List of organisms in the evolutionary profiles for D-GluRS and $tRNA^{Glu}$ (continued).

Class I alignment The Class I structural alignment was prepared by aligning the catalytic domains of aaRS·substrate structures in MultiSeq (1N78 [316], 1QTQ [342], 1F7U [343], 1LI7 [344], 1FFY [345], 1H3N [346], 2CT8 [347], 1GAX [348], 2AKE [298], and 1H3E [349]). Triad residues were located by their proximity and orientation relative to the

α-amino group on the adenylate analog or charging amino acid. Sequence alignments were prepared by downloading all Class I aaRS sequences from the Swiss-Prot database [350] by domain of life and specificity. These were aligned with ClustalW and manually improved. Evolutionary profiles were created using SeqQR with a sequence identity cutoff of 70%, yielding statistically well-balanced sets for domain/specifities with either many sequences (bacterial) or few sequences (eukaryal). The sequence of the aaRS structure in a given specificity was aligned to each evolutionary profile to locate the columns containing triad residues. For columns containing multiple residue types, the residues are listed in Figure 3.8b from most frequent to least and colored by percent sequence identity (blue = most conserved, white = 50% conserved, and red = unconserved). Columns containing more than four residue types in the evolutionary profile are considered unconserved.

3.3.2 Molecular modeling

Pre-transfer complex The Pre-transfer model was based on GluRS·tRNA^{Glu}·analog crystal structure 1N78 [316], which contains *T. thermophilus* GluRS, transcribed tRNA^{Glu}, and adenylate analog glutamol-AMP. The glutamol-AMP was modified to glutamyl-AMP by adding a carbonyl oxygen using the Psfgen plugin in VMD [151]. Histidine protonation states were investigated with the Whatif server [351] and visual inspection.

Post-transfer complexes All post-transfer complexes were modeled from the crystal structure. The backbone carbonyl carbon was detached from the AMP and reconnected to the 2′ ribose oxygen of A76 to create Glu-tRNA^{Glu}. This new bond was 4.00 Å in the model and relaxed to 1.35 Å during minimization. In the Post (H-AMP) state, a proton was attached to the AMP phosphate oxygen that previously had been bonded to the charged glutamate. The capture of the proton by water was modeled in the Post (AMP) state by removing the 2′-OH proton from the system and compensating for the charge difference by adding a monovalent ion to the bulk solution. To test the validity of the Post (H-AMP) and Post (AMP) states, an intermediate state was modeled from the Post

(H-AMP) system after 5 ns of equilibration wherein the proton on the H-AMP was removed and a potassium ion was added to bulk solution. This system was then equilibrated for 20 ns during which active site residues relaxed to positions similar to those in the equilibrated Post (AMP) state (data not shown).

In the Post (AMP/H15h) state, His15 in the HIGH motif was protonated (residue type HSP in the CHARMM27 force field). In the Post (AMP/GluNH$_2$/E41COOH) state, the α-ammonium group was deprotonated and Glu41 was protonated using the glutamic acid patch in CHARMM. In the Post (no AMP/GluNH$_2$) state, a proton from the α-ammonium group on the charged glutamate was removed.

Free tRNA The free tRNA^{Glu} system was modeled from the crystal structure in the absence of GluRS or small molecule substrates. The free charged tRNA^{Glu} was modeled modeled similarly from the Post (AMP) state.

Migration complex The migration complex was assembled from proteins and tRNA in two different simulations. The EF-Tu structure was taken from an EF-Tu·GTP·Cys-tRNA^{Cys} trajectory in [63], and the bound GluRS and Glu-tRNA^{Glu} structures were taken from the Post (no AMP) trajectory. The structures in the first frames of each simulation were overlapped based on a structural alignment of their respective tRNA backbones, excluding the anticodon arms. Then each trajectory was aligned back to its first frame based on the same tRNA backbone selection. The trajectory frames selected contained GluRS and EF-Tu structures that lacked steric clashes when aligned by tRNA backbone. These were merged to create the final migration complex. Water and ions within 10 Å of GluRS·tRNA^{Glu} and EF-Tu were included with the migration complex unless they produced steric clashes with molecules in the system. Finally, the remaining solvent box was generated and neutralizing K^+ atoms were added.

Another system was created based on the migration complex after 10 ns of simulation. A proton from the α-ammonium group on the charged glutamate was removed and one of the bulk water molecules was replaced by a K^+ atom to maintain charge neutrality. After the

next 20 ns, this group was reprotonated, and a K^+ atom was removed from the system.

3.3.3 Molecular dynamics

System setup Psfgen was used to add hydrogen atoms to the molecules. Each system state was neutralized by placing magnesium and potassium ions with ionize [2], a program that calculates the Coulombic interaction energy for the placement of an ion on a uniform grid. Each ion is added to the system at the energetic minimum, and the process is repeated until all ions have been placed. The Mg^{2+} placement protocol developed previously was used to place three Mg^{2+} on the primary solvation shell of the tRNA (at 2 Å) and 14 Mg^{2+} and 44 K^+ at 6.5 Å from the complex in its ionic cloud [63]. For states with changes in charge within the solute, the number of neutralizing K^+ ions was modified accordingly. The same number protocol was used for the free tRNA systems with the number of K^+ ions decreased accordingly. The Mg^{2+} ion primary solvation shells were filled with a maximum of six TIP3 water molecules [352].

The concentration and motion of Mg^{2+} ions have powerful effects on nucleic acid structure and dynamics (for a review see [353]). As demonstrated previously with EF-Tu·Cys-$tRNA^{Cys}$ studies [63], Mg^{2+} ions disrupt RNA backbone conformations leading to structural fluctuations when they move in and out of tRNA hairpin loops. However, by interacting with the RNA deep groove, cations can stabilize helices by shielding the repulsive electrostatic interactions between phosphates on either side of the groove. An RNA helix with Mg^{2+} ions in its deep groove becomes more rigid and less able to conform to an induced fit at a protein binding interface. In this study, several replicate simulations were performed for each system state, and the Mg^{2+} ion distribution after 5 ns was assessed through ion occupancy and residency calculations to determine if there were too many (> 8) or too few (< 6) resident Mg^{2+} atoms associated with the tRNA. This check ensured that the initial movement of Mg^{2+} ions was less likely to lead to unusual structural fluctuations. The residency locations were determined by tracking the positions of the

[2] http://www.ks.uiuc.edu/Development/MDTools/ionize

Mg^{2+} ions over time. Frames from the last 16 ns of each simulation (40 ps intervals) were aligned by the tRNA backbone. Mg^{2+} ions remaining within 3 Å of a starting position for 4 ns were classified as resident ions. Occupancy maps were calculated using the VolMap plugin to VMD with 1 Å resolution and averaged over all frames.

Solvate 1.0[3] was used with two gaussians to add two layers of water molecules to the system. This resulted in placement of eight additional water molecules in the active site of the Pre-transfer system. The Solvate 1.2 plugin in VMD was used to place the bulk water, reaching a an average system size of approximately 120 x 80 x 120 Å and 110,436 atoms.

The NAMD2 software [354] and CHARMM27 force field [355] were used to perform the MD simulations. The simulations were calculated in the NPT ensemble with periodic boundary conditions; the Langevin piston method [356] was applied to maintain pressure at 1.01325 bar and Langevin dynamics were used to maintain temperature at 298.15 K. Electrostatics were efficiently treated with the Particle Mesh Ewald summation [357]. A multiple time-stepping algorithm was used to evaluate bonded interactions at 1 fs, vdW interactions every 2 fs, and electrostatic forces every 4 fs. A cutoff of 12 Å and a switching distance of 10 Å were used for the vdW force calculations.

Minimization Minimization was carried out according to the protocol previously established for protein·RNA complexes [63] to ensure that water molecules were associated with the macromolecules prior to removing all constraints on macromolecule motion. The system was minimized in four stages: 2,000 steps with heavy atoms fixed, 3,000 steps with heavy atoms fixed excluding water, 5,000 steps with macromolecule backbone atoms fixed, and 20,000 steps of unconstrained minimization.

Equilibration Equilibration was conducted using a temperature jump protocol similar to that developed by Auffinger and Westhof [358]. This enabled cations to enter the deep groove of the negatively charged nucleic acid by systematically raising the temperature and freeing harmonic constraints in five steps [63]. These steps were: a temperature of 100 K

[3]Grubmueller, H. 1996 Solvate 1.0. http://www.mpibpc.gwdg.de/abteilungen/071/solvate/docu.html

with heavy atoms and ions constrained (25,000 fs), 200 K with hydrogens and waters freed (25,000 fs), 250 K with all hydrogens, waters, ions, and small molecules unconstrained (25,000 fs), 250 K with macromolecule backbones constrained (25,000 fs), and 298.15 K with no constraints (19.9 ns). The harmonic constraints in all steps were $1\,\mathrm{kcal\,mol^{-1}\,\AA^2}$.

Parameters Parameters for glutamyl-AMP were derived by analogy to AMP and glutamate in the CHARMM27 force field. Parameters for the ester bond connecting the charging glutamate to the $2'$ hydroxyl group of A76 were analogous to those developed previously from the ester bond in fatty acids [63]. H-AMP parameters were derived by analogy to pyrophosphate, protonated pyrophosphate, and AMP. Parameters for the deprotonated α-amino group on the charging glutamate were the same as those from [63].

3.3.4 RMSD calculations

The average RMSD for the protein and nucleic acid was calculated separately using the C^α and phosphorous atoms at 40 ps intervals over the 20 ns simulations. The C^α and phosphorous atoms of the complex backbone in the crystal structure were used as the reference for the standard Kabsch method [359] implemented in VMD.

3.3.5 pKa calculations

Protonation states for titratable protein and ligand groups in the active site were assigned using PROPKA 2.0 [360]. PROPKA uses a fast heuristic method to compute pKa perturbations of the titratable sidechains due to desolvation, hydrogen bonding, and charge-charge interactions in the environment of the protein. A well equilibrated frame at the end of the 20 ns of simulation for each state was used in the pKa calculation. Bound magnesium ions in the primary solvation shell of the tRNA were found to have minimal effect on the pKa of the protein sidechains and were therefore not included. The bulk pKa values were set to 6.9 for the AMP/H-AMP phosphate and 9.47 for the amino acid moiety

on the AMP adenylate and A76. Other chemical groups on the AMP, adenylate, and A76 were included using the PROPKA default representations for those groups.

3.3.6 Principal component analysis

Collective motion of GluRS and $tRNA^{Glu}$ in differing system states was investigated using a standard principal component analysis based on the C^{α} and P atoms in the molecular backbones. The unnormalized covariance matrix *Cov* was obtained by averaging over the last 5 ns (10,000 frames) of each trajectory (15 to 20 ns in all trajectories except the Post (no AMP/$GluNH_2$) at 75 to 80 ns). $Cov_{ij} = \langle \Delta\vec{r}_i(t) \cdot \Delta\vec{r}_j(t) \rangle$ where $\Delta\vec{r}_i(t) = \vec{r}_i(t) - \langle \vec{r}_j(t) \rangle$, and $\vec{r}_i(t)$ is the position vector of the C^{α} or P atom of the i^{th} residue or nucleotide, at time t. The covariance matrix, eigenvalues, and eigenvectors were calculated with the program CARMA [361]. The contribution of each eigenvector to the full system motion is obtained from the projection matrix and is shown in Figure 3.3 for the Pre-transfer and Post (no AMP/$GluNH_2$) states and Figure 3.4 for the free $tRNA^{Glu}$ and Glu-$tRNA^{Glu}$ simulations. The data for each principal component was projected onto the system and the RMSD of each residue was calculated relative to the first frame of the transformed trajectory.

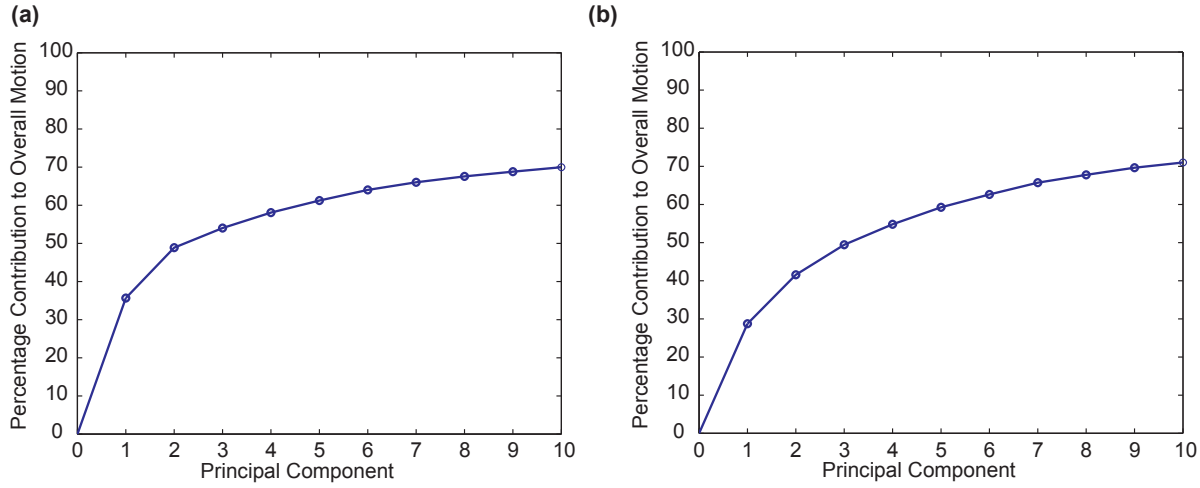


Figure 3.3: Comparison of the principal component percentages for the last 5 ns of the Pre-transfer and Post (no AMP/$GluNH_2$) system trajectories.

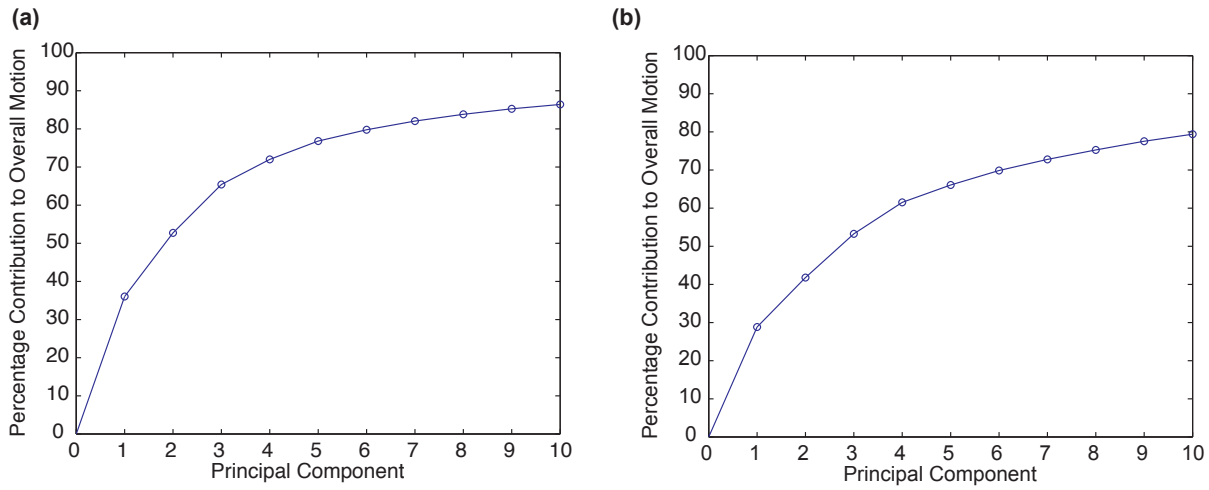


Figure 3.4: Comparison of the principal component percentages for the last 5 ns of the free $tRNA^{Glu}$ and Glu-$tRNA^{Glu}$ system trajectories.

3.3.7 Correlation calculations

To calculate correlations within and between GluRS residues and $tRNA^{Glu}$ nucleotides, the backbone atoms from the last 5 ns (10,000 frames) of each 20 ns trajectory were used to compute the normalized covariance (correlation) matrix with Carma. The elements of this matrix are the correlations in motion of the residues and nucleotides. Values are close to one if two residues/nucleotides are moving in the same direction in a majority of frames (correlation), approaching negative one if they move in opposite directions (anticorrelation), and near zero if their motion is uncorrelated.

3.3.8 Dynamical network construction and community analysis

A weighted GluRS·$tRNA^{Glu}$ network was constructed for each state based on dynamical contacts as outlined previously [131]. Each amino acid, nucleotide, or small molecule substrate/product is represented by a node. Edges between nodes are defined by dynamical contacts, which are present when the heavy atoms of two residues/nucleotides are within 4.5 Å of each other in 75% of the trajectory frames analyzed (50 ps intervals from the last 5 ns of each simulation). Covalently bonded neighbors were neglected. The edge weights

w_{ij} were calculated from the correlation values (C_{ij}) which define the probability of information transfer across each edge: $w_{ij} = -log(|C_{ij}|)$.

Shortest paths through the network represent the dominant mode of communication between two nodes. Path lengths are the sum of the edge weights between consecutive nodes (k,l) along the path, $D_{ij} = \Sigma_{k,l} w_{kl}$, and the shortest paths were determined using the Floyd-Warshall algorithm [362, 363]. Betweenness of an edge is defined as the number of shortest paths traversing that edge, and high betweenness values identify contacts that are important for communication across the complex.

The Girvan-Newman algorithm was used to identify tightly interconnected modules or communities of nodes [364]. This algorithm iteratively removes edges with the highest betweenness values and recalculates the betweenness of all remaining edges to determine the optimal community structure as measured by the modularity score. This score is a measure of the difference between the probability of inter- and intracommunity edges and is maximized during the calculation.

3.3.9 Local energetics analysis

Nonbonded energetics were calculated using the NAMDEnergy plugin in VMD with the CHARMM27 force field. VdW and electrostatic interaction energies were calculated between each protein residue and either the tRNA (omitting the charging glutamate) or the charging amino acid. Energies were calculated at 10 ps intervals and averaged over the last 5 ns of each trajectory. The switching and cutoff distances of 18 Å and 21 Å ensured that electrostatic energy was calculated relative to both members of a nucleic acid base pair. To determine if the residues with strong interaction energies were conserved, a mask was calculated based on the multiple sequence alignment (see Bioinformatics in Methods). The average interaction energies per residue were scaled by percent sequence identity over the bacterial sequences based on the *T. thermophilus* sequence.

3.3.10 MM-PBSA Calculation of Free Energies

The Molecular Mechanics–Poisson-Boltzmann Surface Area (MM-PBSA) method is used to calculate binding free energies from a simple thermodynamic cycle [365, 366, 367, 368, 369, 370, 63, 371]. The difference in energy between the complex and two components or unbound docking partners is calculated according to $\langle \Delta G_{binding} \rangle = \langle \Delta E_{VdW} + \Delta E_{coul} + \Delta G_{nonpolar} + \Delta G_{polar} \rangle - T\langle \Delta S \rangle$. Each value in the averaged terms is the difference for either a small molecule substrate/product undocking from the protein·tRNA complex ($\langle \Delta G^{AMP}_{binding} \rangle = \langle G_{complex} - G_{GluRS\cdot tRNA} - G_{AMP} \rangle$), or tRNA dissociating from the GluRS·substrate/product complex ($\langle \Delta G^{tRNA}_{binding} \rangle = \langle G_{complex} - G_{GluRS\cdot AMP} - G_{tRNA} \rangle$). Three bound Mg^{2+} ions placed during system setup and water molecules in their first solvation shell were included with the tRNA, all of which bind the tRNA core and are stable elements of the nucleic acid structure. Values were averaged over 500 frames from the final 5 ns of each 20 ns simulation (10,000 frames for the entropy calculation). Calculations for individual components were performed using the same trajectory as the complex [367, 63], neglecting contributions from conformational changes that could occur were the individual components simulated separately.

The bonded interactions (bonds, angles, dihedrals, and impropers) cancel in each difference calculation and therefore are not shown. The E_{VdW} term is the sum of all pairwise VdW interaction energies and was calculated using the NamdEnergy plugin in VMD. The Coulombic energy (ΔE_{coul}), which is the sum of all pairwise interactions in the system scaled by ϵ_{in}, was calculated with Coulomb, a program distributed with APBS [372].

The nonpolar solvation free energy ($\Delta G_{nonpolar}$) is the energetic cost of creating a cavity in the solvent for a given system. $G_{nonpolar} = \gamma SASA + b$ was calculated for the complex and components with SASA set to the solvent accessible surface area using a solvent radius of 1.4 Å, $\gamma = 0.00542\,\text{kcal mol}^{-1}\ \text{Å}^{-2}$, and $b = 0.92\,\text{kcal/mol}$. This term was calculated in VMD. The polar solvation (ΔG_{polar}), the energy required to move the charged solute from a vacuum dielectric ($\epsilon_{in} = 1.00$) to an aqueous dielectric ($\epsilon_{out} = 78.54$), was computed for the complex and individual components. The 1.00 interior dielectric was chosen to make the

free energy of the Pre-transfer state approximately -8 kcal/mol, consistent with experiment (see Results). This term was numerically calculated with the Poisson-Boltzmann solver APBS [372]. A temperature of 298.15 K and CHARMM27 radii/charges were used on a grid of 193 x 129 x 193 with automated focussing. Grid size was based on the crystal structure with 15% padding in each dimension. The implicit salt concentrations of 117 mM KCl and 38 mM $MgCl_2$ were used to model the explicit concentrations in the system states. Each frame was aligned to the crystal structure based on the C^{α} and P atoms to limit error from orientation changes. The grid was centered on the complex.

The entropy terms (ΔS) were estimated from the covariance matrix of atomic position fluctuation using Schlitter's formula [373, 374, 375], which relates the entropy of a solute molecule to the sum of decoupled simple harmonic oscillators obtained from the principal component modes in the molecular dynamics simulation. The coordinates of the backbone atoms (C^{α} and phosphorous) in 10,000 frames from the last 5 ns of each trajectory were used to generate the mass-weighted covariance matrix; Carma was used to diagonalize it. The resulting eigenvalues were then used to calculate the determinant and substituted into Schlitter's formula to calculate the entropy of the solute [63]. Because the small molecule substrate/products had only one or two backbone atoms, which is an insufficient number to calculate the covariance matrix, the entropy was estimated using the Sakur-Tetrode equation for dimolecular and unimolecular ideal gases [376].

3.4 Results and Discussion

3.4.1 Charging mechanism and post-transfer states

A concerted mechanism has been proposed for tRNA aminoacylation in the homologous GlnRS·$tRNA^{Gln}$ system [317] which was later experimentally tested on the Class I TyrRS·$tRNA^{Tyr}$ complex [377]. In the transfer step of the proposed aminoacylation reaction mechanism, the 2′-hydroxyl oxygen from A76 nucleophilically attacks the α-carbonyl carbon of Glu-AMP while, simultaneously, the A76 2′-hydroxyl proton shifts to

the Glu-AMP α-phosphate, resulting in Glu-tRNA^{Glu} and H-AMP [317]. Assuming a similar concerted mechanism holds for GluRS·tRNA^{Glu}, we created *in silico* a series of relevant post-transfer states (see Methods). The states were designed to answer three questions regards regions of the system that could change. First, we assume that the Post (H-AMP) state exists. Since the proton on H-AMP is not expected to be stable, is the proton transferred to water or a nearby active site residue that could act as a general base? His15 in the HIGH motif is nearby and therefore is a candidate. We therefore modeled the Post (AMP) and Post (AMP/H15h) states to determine effect of proton transfer to either water or His15 on the active site. Second, following transfer of the charging amino acid to tRNA, the AMP has no further function. Does the AMP dissociate from the active site while the tRNA remains docked and could the dissociation affect tRNA binding? We modeled the Post (no AMP) state to determine if lack of AMP would affect the active site. Finally, the α–ammonium group on the charging glutamate is known to be in the charged state due to the three highly conserved active site residues making hydrogen bonds and salt bridges to the three protons. If a proton was transferred and the α–amino group was neutral, one interaction would be lost and this might spur dissociation. The Post (no AMP/GluNH_2) state was modeled to determine the effect of removing AMP and neutralization of the α–amino group on the charging glutamate. Figure 3.2 provides a summary of the terminology and differences in initial active site modeling between the simulations.

3.4.2 Modeling Pre- and Post-transfer states

Pre-transfer state The adenylate analog in the GluRS·tRNA^{Glu} crystal structure (PDB code 1N78) is chemically inert because it lacks the α-carbonyl group on the glutamate backbone. Upon replacing the analog with Glu-AMP and equilibrating the system, small rearrangements occur around the α-carbonyl of the glutamate moiety and phosphate of the AMP moiety. The 2′-hydroxyl of A76 reorients towards the α-carbonyl group (see Figure 3.6), positioning the reactants with the geometry required for aminoacylation. To

accommodate these changes, the distance between the 3′-hydroxyl of A76 and the α-ammonium group on the charging glutamate lengthens. Contact distances between highly conserved Ala7, Ser9, and Glu41 and the α-ammonium group shorten. Contacts between the charging glutamate sidechain and other active site residues lengthen slightly as the system relaxes. The final configuration of the active site was within 2 Å of the crystal structure, demonstrating that the initial protonation states and modification of the analog produced minor perturbations (see Figure 3.5). The MD simulation of the Pre-transfer state was therefore used as a control, providing the baseline behavior of the system.

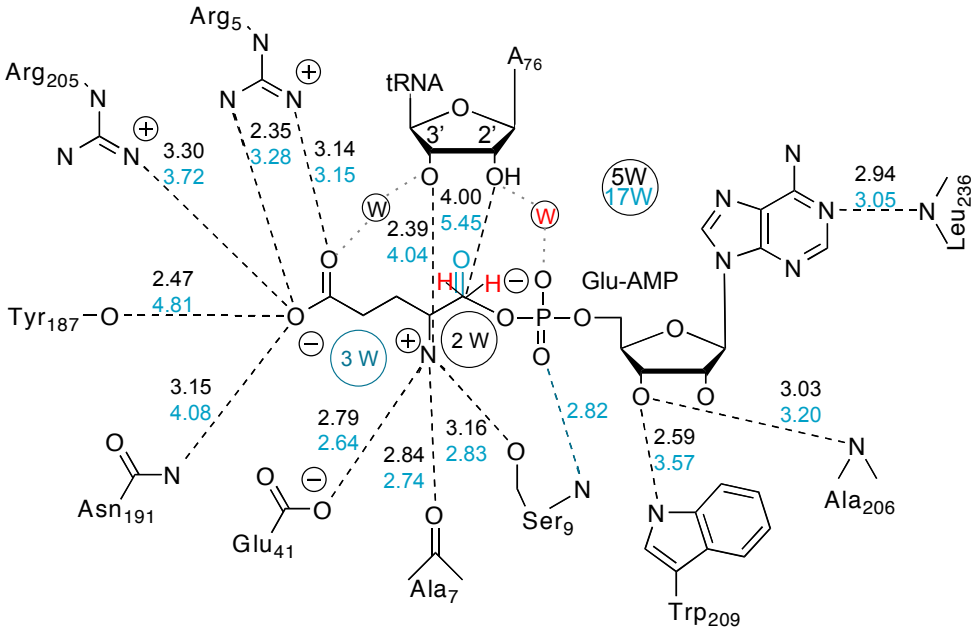


Figure 3.5: Comparison of the active site between the crystal structure and the equilibrated Pre-transfer state. Distances are measured between the heavy atoms with black values from the crystal structure and blue values being averages over the last 16 ns of the 20 ns Pre-transfer trajectory. Atoms and bonds in blue are unique to the trajectory. The CH_2 in the analog and a crystal water are colored red. Circled "W"'s indicate regions of high water density with gray bonds indicating waters mediated contacts.

Assignment of the protonation state of residues at the GluRS·tRNA^{Glu} interface and within the active site is a key consideration when developing models for protein·RNA dissociation. The pKa values of all titratable sidechains on GluRS in the Pre-transfer state

were estimated at the end of the simulation (see Methods), and the values for residues within the active site are reported in Table 3.2. In the Pre-transfer state, all histidine residues (including His15 in the HIGH motif) are predicted to have pKa < 7, indicating that the neutral form for each histidine is dominant at pH 7.0. All lysine/arginine residues are predicted to have pKa > 7 while the acidic aspartate/glutamate residues are predicted to have pKa < 7, suggesting that they exist as charged residues in the Pre-transfer state.

	pKa Calculations after 20 ns Simulations				
Residue	Pre-transfer	Post (HAMP)	Post (AMP)	Post (AMP/$GluNH_2$ /E41COOH)	Post (no AMP/ $GluNH_2$
(H-)AMP phosphate	**-**	**5.51**	**7.70**	**6.44**	**-**
Glu amino	**9.86**	**7.23**	**7.69**	**5.12**	**9.47**
Glu carboxy	2.11	0.89	2.50	0.38	4.50
Arg5 sidechain	11.91	11.09	10.35	10.07	9.29
His15 sidechain	2.79	3.52	2.93	3.48	2.24
Glu41 sidechain	**1.86**	**8.06**	**8.90**	**4.50**	**6.03**
Tyr187 sidechain	13.11	12.08	11.18	11.39	15.75
Arg205 sidechain	11.45	9.54	11.15	12.27	12.98
Glu208 sidechain	3.94	-4.91	3.39	2.99	3.25
Lys243 sidechain	11.10	10.94	11.18	11.03	9.76

Table 3.2: Predicted pKa values for residues within 5 Å of the small molecule substrate/products, the (H-)AMP, and the α–amino group of the charging glutamate in the Pre-transfer state and several post-transfer states. Residues with pKa values changing from above or below 7 across the states are shown in bold. All states were run at pH 7.

In the first post-transfer state (Post (H-AMP)), the amino acid moiety on the adenylate is transferred to the tRNA, and the proton previously on the 2′-OH of A76 is transferred to the AMP phosphate, which is the only general base in close proximity to the ribose hydroxyl group. As this state models the system immediately following the transition state, it is expected to exhibit binding affinities similar to the Pre-transfer state. Separation of H-AMP from the amino acid moiety causes a few significant rearrangements

in the active site (see Figure 3.6b and c). The H-AMP α-phosphate shifts closer to the HIGH motif and forms hydrogen bonds with His15 and Thr18. Both residue sidechains rotate to accommodate the new interactions. In bulk solvent, H-AMP has a pKa of 6.9 [378], but in the protein, interactions with His15 and Ser9 reduce the pKa of the H-AMP phosphate to 5.51 (see Table 3.2). This suggests that the proton could be further transferred to another residue on the protein or to a water molecule, forming the Post (AMP) state. Because its sidechain is buried within the protein, neutral His15 is predicted to have a pKa of 3.52 and therefore is unlikely to serve as a general base for abstracting the H-AMP proton. For confirmation, a system containing AMP and protonated His15 was simulated (Post (AMP/H15h) state), but the free energy analysis showed that this state has tighter binding at the complex interface than the Pre-transfer state (see Table 3.9) and therefore does not lead to tRNA dissociation.

The electrostatic and hydrogen bonding environment around the Glu41 sidechain causes its local pKa to rise to 8.06 at the end of the Post (H-AMP) simulation and 8.90 after 20 ns in the Post (AMP) state (see Table 3.2). Because these simulations were performed at pH 7.0, Glu41 most likely exists in the neutral glutamic acid form in the Post (AMP) state. The α-ammonium group of the charging glutamate has a predicted pKa of 7.23 and therefore can be in either the cationic or neutral form. This suggests that the nearby negatively charged Glu41 sidechain might deprotonate the α-ammonium group and participate in further proton transfer events. A Post (AMP/GluNH_2/E41COOH) simulation, similar to the Post (AMP) state in all respects except that a proton from the α-ammonium group was transferred to the Glu41 sidechain resulting in a neutral α-amino group and a glutamic acid, was also run. The pKa values at the end of this simulation suggest that the neutral Glu41 (4.50) becomes deprotonated after exposure to solvent, but the α-amino group (5.12) remains neutral.

The next set of states was modeled under the assumption that AMP had already undocked. Initially, eight water molecules were placed in the vacant AMP binding site. This site becomes further hydrated as the increasingly mobile KMSK loop allows water

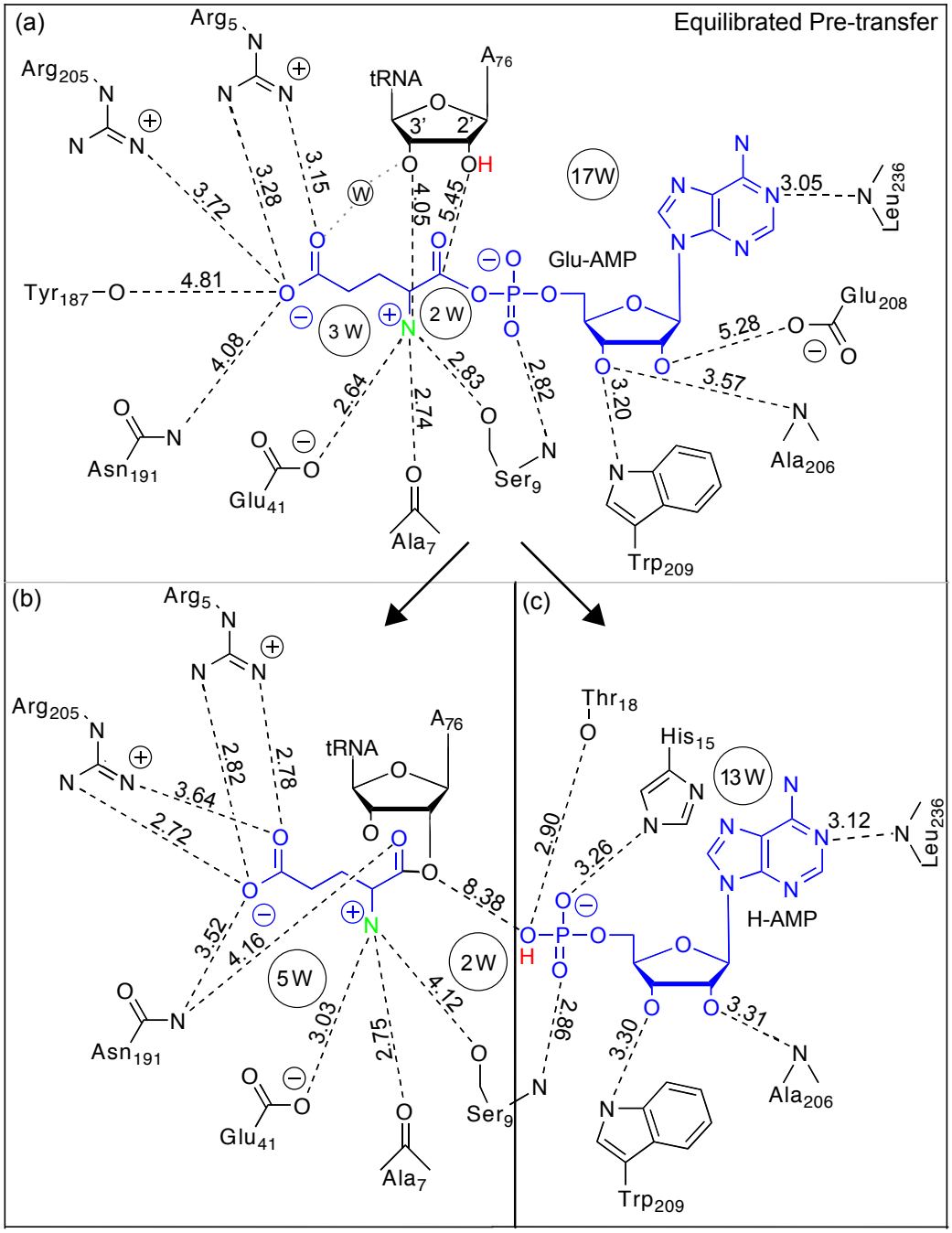


Figure 3.6: Interactions between the small molecule substrate/product with GluRS active site residues. In (a) Glu-AMP interacts with A76 on the tRNA, residue sidechains, and water. Panels (b) and (c) show the residues around the (b) charging glutamate and (c) H-AMP in the equilibrated Post (H-AMP) state. The Glu-AMP, charging glutamate, and H-AMP atoms are colored blue while the H-AMP proton is colored red and the glutamate α–nitrogen is colored green for emphasis. All hydrogen atoms other than the H-AMP proton have been removed for clarity. The distances shown were measured between the heavy atoms and averaged over the last 16 ns of the 20 ns trajectories.

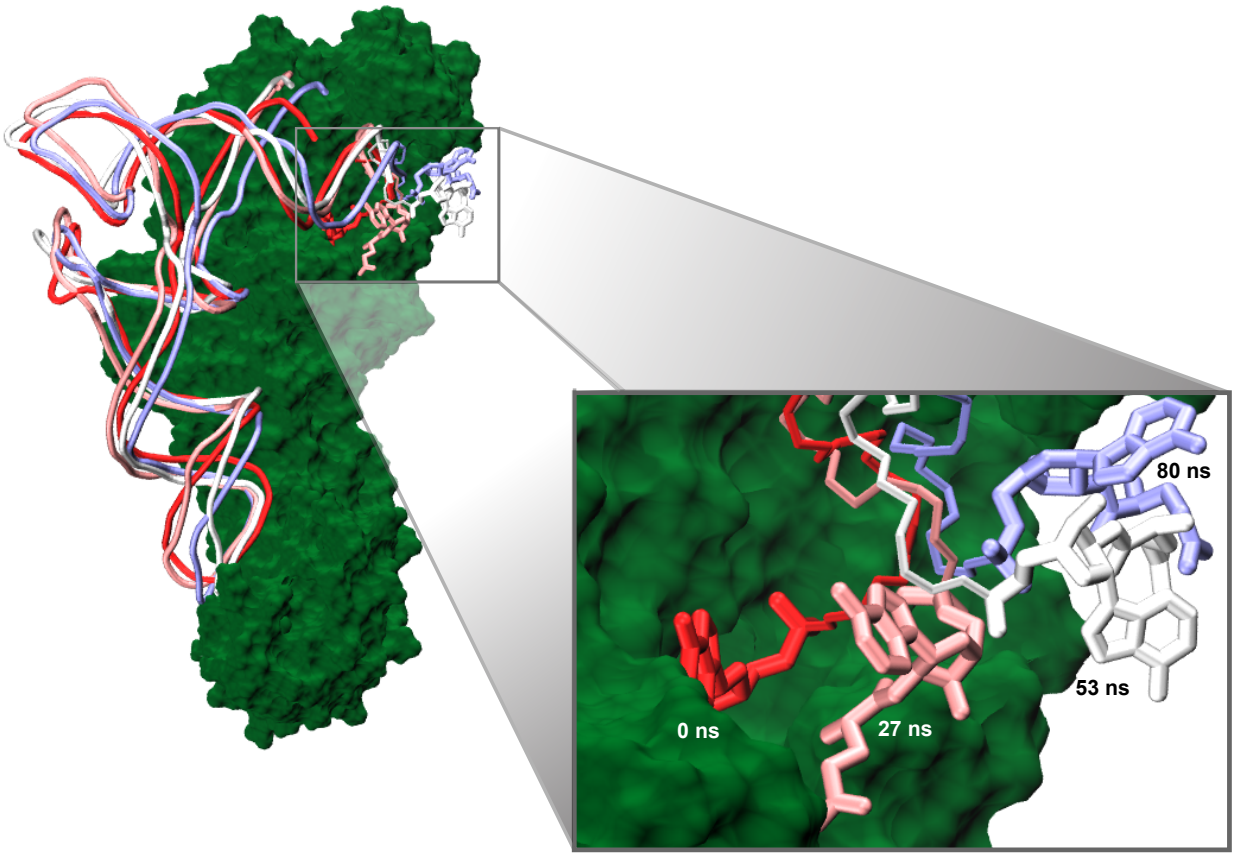
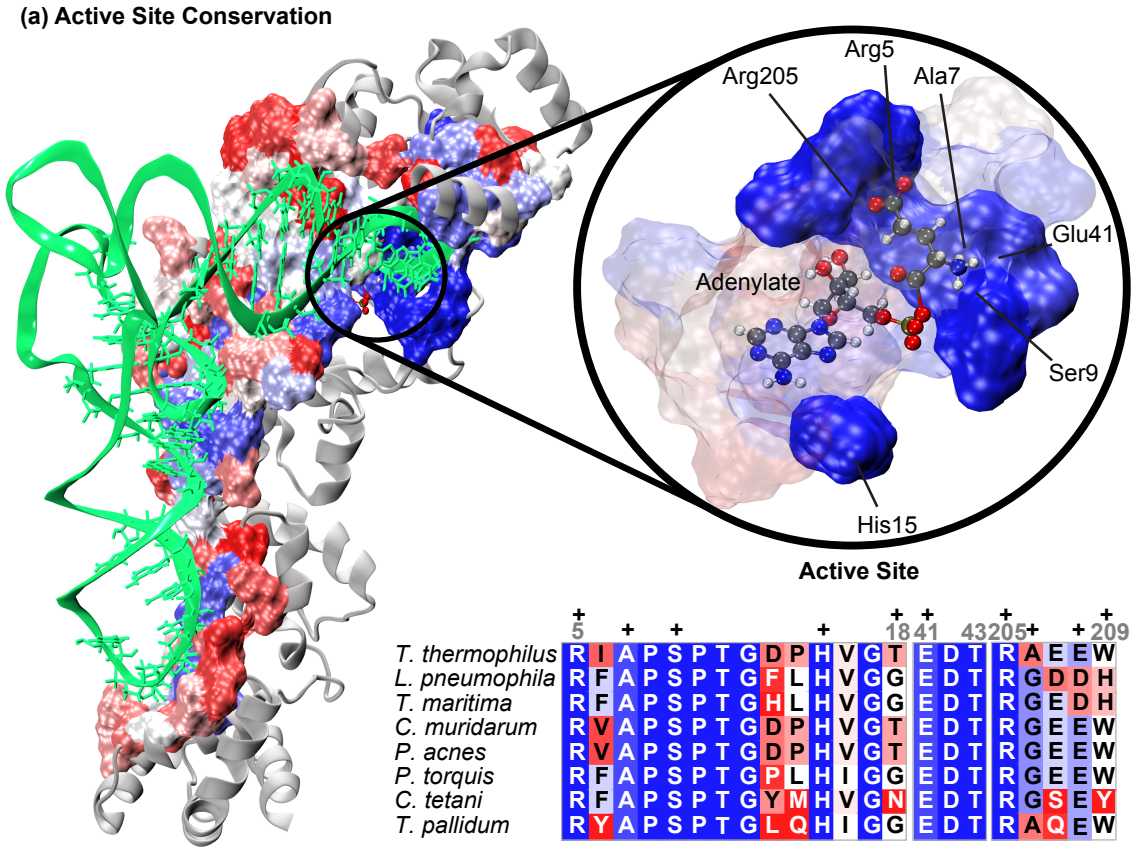


Figure 3.7: Motion of tRNA over 80 ns in the Post (no AMP/GluNH$_2$) state. The tRNA and charging amino acid are shown at four evenly distributed timepoints along the trajectory. Each frame was aligned to the crystal structure by the protein.

molecules to enter. These waters penetrate deep into the active site, weakening contacts between the charging amino acid and multiple active site residues. Simulations of the Post (no AMP/GluNH$_2$) state, in which the Glu41 sidechain is charged again and the α–amine of the charging glutamate is neutral, are expected to show fast release of the tRNA as the water molecules that replace the AMP moiety assist in solvating and weakening the interactions between the charged tRNA and GluRS. The interactions of the charging glutamate sidechain with Arg5 and Arg205 are solvated, leading to greater motion that loosens the contacts between the α–ammonium group and its triad of binding residues. Once the charging amino acid breaks its contact with the Glu41 sidechain in the Post (no

AMP/GluNH_2) state, the pKa of the α-amino group increases to 9.47 (see Table 3.2). This indicates that the α-amine of the charging amino acid is likely to be reprotonated while the Glu41 sidechain displays a propensity to remain charged upon losing contact. The motion of the CCA hairpin as it undocks over 80 ns is shown in Figure 3.7.

Universal handle for charged tRNA release Glu41 binds the α–ammonium group, which is a fundamental group on all charging amino acids. Figure 3.8b shows that Glu41 (amino acid 3) is conserved as either glutamate or aspartate in GluRS enzymes across all domains of life. Further, comparison of ten available crystal structures of Class I aaRSs [316, 342, 343, 344, 345, 346, 347, 348, 298, 349] reveals that in nearly all specificities a conserved aspartate or glutamate is in the same orientation in the active site and binds the α–ammonium group of the charging amino acid, confirming previous studies [320]. This residue is analogous to the "universal aspartate" found in the editing domain of LeuRS and IleRS that interacts with the same α–ammonium group [379, 380]. The pKa calculations and sequence conservation studies suggest that Glu41 may represent a universal handle in Class I aaRSs to stabilize the charging amino acid during aminoacylation. The transfer of a proton from the α–ammonium group to the universal handle may be a standard strategy for facilitating release of the charged tRNA following the transfer reaction. The other residues forming the conserved triad binding the α–ammonium group are usually conserved in structure but not in sequence over all Class I aaRSs. They are conserved in sequence within each specificity over all three domains of life, possibly indicating coevolution with the charging amino acid. The first (amino acid 1) includes Ala7 in bacterial GluRS and commonly binds the α–ammonium group through its main chain carbonyl oxygen. This is analogous to the methionine in the LeuRS editing domain that stabilizes the charging amino acid for hydrolysis through a hydrogen bond between the methionine main chain carbonyl oxygen and the charging amino acid α–amino group [380]. The second (amino acid 2) includes Ser9 and generally forms a hydrogen bond through the sidechain. This residue, which is a glutamine in TyrRS, was previously shown to stabilize the transition state of the second step of aminoacylation [377]. Interestingly,



(b) Triad Conservation

aaRS (PDB)	Organism in Structure	Amino Acid 1				Amino Acid 2				Amino Acid 3			
		Str	B	A	E	Str	B	A	E	Str	B	A	E
Glu (1n78)	*T. thermophilus*	A7	ACP	A	AP	S9	SNE	NS	SE	E41	ED	ED	ED
Gln (1qtq)	*E. coli*	P32	P	-	P	E34	ED	-	E	D66	D	-	D
Arg (1f7u)	*S. cerevisiae*	S151	SGA	S	S	N153	N	N	N	E294	-	-	E
Cys (1li7)	*E. coli*	G29	G	G	G	T31	T	T	T	T68	T	T	T
Ile (1ffy)	*S. aureus*	P56	P	P	P	Y58	-	YT	FY	D95	D	D	D
Leu (1h3n)	*T. thermophilus*	F41	FLY	-	YFL	Y43	Y	Y	Y	D80	DH	HDQ	HD
Met (2ct8)	*A. aeolicus*	I13	-	LW	L	Y15	Y	Y	Y	D52	D	D	D
Val (1gax)	*T. thermophilus*	P42	P	P	P	N44	NTY	YTF	N	D81	D	D	D
Trp (2ake)	*H. sapiens*	Q313	QN	Q	Q	Q284	MQ	Q	QE	E199	HQ	E	E
Tyr (1h3e)	*T. thermophilus*	Y175	Y	Y	Y	Q179	Q	Q	Q	D85	D	-	-

Figure 3.8: Conservation of GluRS at the tRNA binding interface and active site. In (a), residues within 5 Å of tRNA are colored by percent sequence identity. Blue, white, and red indicate strong, moderate, and poor conservation, respectively. On the right, highly conserved active site residues active site make contacts with the adenylate in solid blue with nearby residues as transparent. A subset of the D-GluRS alignment is shown, representing major bacterial phyla. '+' residues interact with the adenylate either here or in Figure 3.6. In (b), structures in the Class I alignment are shown along with conserved triad residues (in order of occurrence in GluRS) for each particular structure. Adjacent columns list the residues in the sequence alignment for each domain of life (B = *Bacteria*, A = *Archaea*, and E = *Eucarya*). Residues not conserved within a domain of life are denoted by a dash.

IleRS, LeuRS, and MetRS each contain an aromatic residue in the same structural position, which does not form a hydrogen bond, but instead stacks on the backbone of the charging amino acid. This difference may be necessary to stabilize these hydrophobic charging amino acids.

The conservation of amino acid 3 in both sequence and structure is nearly universal in Class I aaRSs, with three exceptions. Based on the crystal structure, CysRS uses a threonine to form a hydrogen bond rather than an aspartate for a salt bridge. The close proximity of a zinc ion in the CysRS active site changes the electrostatics to control binding of the sidechain of the charging cysteine. The CysRS·tRNA^{Cys}·Cys-AMP complex was modeled and simulated, and after equilibration, the charging cysteine α–amino group remained in contact with threonine, not interacting with the nearby aspartate (see Chapter 4). Bacterial TyrRS contains the aspartate in a domain specific insertion, but recent crystal structures of the archaeal and eukaryal TyrRS show that amino acid 3 may be a conserved glutamine in these domains of life [60, 381]. Experimental studies are needed for verification. The bacterial/archaeal TyrRSs are known to be structurally divergent [64]. The aspartate in ArgRS is in a poorly conserved part of the structure, which is due to the wide variety of ArgRS subtypes within each domain of life [64]. Additional crystal structures containing appropriate substrates are needed to provide clarification.

The RMSD of the macromolecular backbones (C^{α} and P atoms) relative to the starting structure was used to monitor the equilibration of each state. There is an initial increase in the protein RMSD that levels off at 1.7 Å after 4 ns and remains approximately constant throughout all the runs. The RMSD of the entire tRNA in the Pre-transfer state remains around 3 Å with the largest fluctuations occuring in the anticodon arm. The motion of the CCA hairpin was monitored by measuring the distance between the charging amino acid and active site residues (Figure 3.9), and it shows the large structural change in the Post (no AMP/GluNH_2) simulation as it undocks.

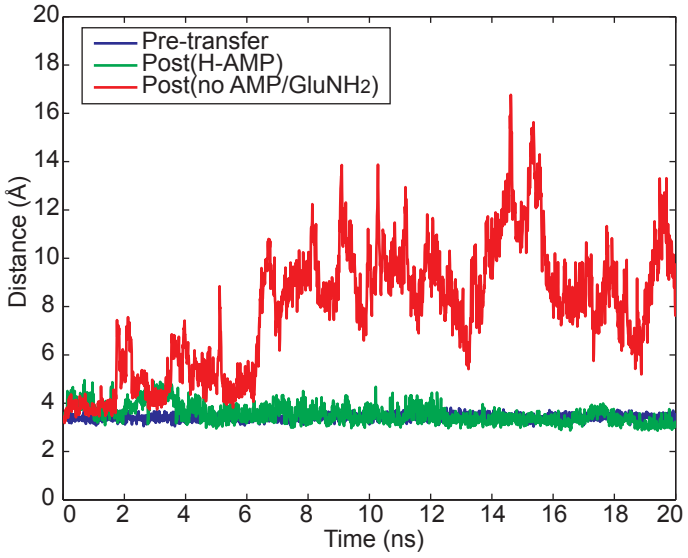


Figure 3.9: tRNA motion in the active site. The distance between the α–amino group on the charging amino acid to the Glu41 sidechain in the active site is shown as a function of time in the Pre-transfer, Post (H-AMP), and Post (no AMP/$GluNH_2$) states.

Magnesium ion perturbations Inclusion of Mg^{2+} ions in the simulations helps capture more physiologically realistic states but also introduces structural perturbations. These random fluctuations are independent of the active site modifications and can hinder the identification of resulting system changes, complicating the comparison of structural and free energy results across different system states. To maintain comparable distribution of Mg^{2+} across the various simulations, a metric was developed using Mg^{2+} ion residency to identify runs in which there were substantial deviations. A resident ion was defined as deviating less than 3 Å from its location for at least 4 ns (see Methods). A comparison of two Pre-transfer runs provides an illustrative example of how different distributions of Mg^{2+} ions cause large fluctuations in tRNA structure. In a replicate run, the complex interface showed substantially weaker interactions near the tRNA acceptor stem and stronger binding near the anticodon arm compared to the first run. A Mg^{2+} ion had entered the deep groove of the acceptor stem and brought phosphates on either side of the groove closer together. The tightening of this helix pulled the first two base pairs of the acceptor stem away from the catalytic domain, resulting in a tRNA conformation that was

significantly different from the crystal structure (the frame of reference used for comparing the different system states). Although this new conformation could be physiologically realistic, the equilibration time required to return the tRNA to a conformation comparable to the crystal structure would have been too computationally expensive. The first run featured seven resident Mg^{2+} ions while the second had nine with the additional two Mg^{2+} ions near the acceptor stem and anticodon arm (see Figure 3.10). To ensure that the system states had similar ion behavior, multiple replicates were performed, and simulations with approximately seven resident Mg^{2+} ions were accepted for further analysis.

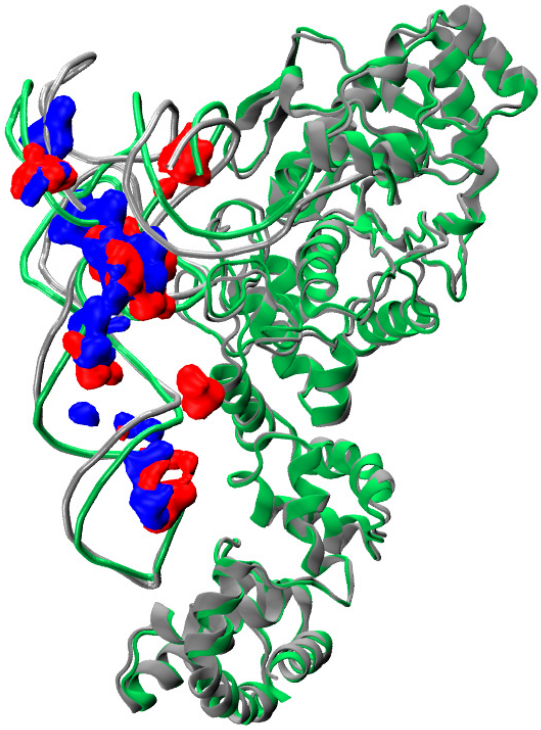


Figure 3.10: Residency of Mg^{2+} ions in two Pre-transfer states. The first run has green tRNA and GluRS with blue occupancy isosurfaces while the second run has gray tRNA and GluRS with red occupany isosurfaces. Isosurfaces mark locations occupied by resident ions.

3.4.3 Charged tRNA has more correlated motion than uncharged tRNA

The slow dissociation of the post-transfer states compared to the Pre-transfer state was measured using standard principal component analysis (PCA). PCA is used to transform a large number of correlated values into a smaller number of uncorrelated values or principal components, and is used here to reveal coupled motion between different regions of a molecule. The covariance of the motion is accounted for in many components with the first representing the largest amount of motion (component 1 contains more motion than component 10). As seen in Figure 3.3, the first 10 principal components account for approximately 70% of the complex motion observed in the last 5 ns of the 20 and 80 ns simulations of the Pre-transfer and Post (no AMP/$GluNH_2$) state respectively. RMSD per residue over the 5 ns trajectory containing the projected motion from one component shows those regions of the molecule that have coupled motion. In Figure 3.11, the RMSD for GluRS and $tRNA^{Glu}$ for the first three components is shown for the Pre-transfer and Post (no AMP/$GluNH_2$) states. GluRS in the Post (no AMP/$GluNH_2$) state has greater motion in the CP1 insertion by a factor of 5 and greater motion in the anticodon binding domain by a factor of 2. These are the two key regions of tRNA binding with the CP1 loop folding over the docked tRNA CCA hairpin and the ACB domain binding the tRNA anticodon. The difference in the tRNA region motion between the simulations is more difficult to interpret due to the small structure of the tRNA, but Figure 3.11b and c show that the largest motions are are split among the first and second components when uncharged and mostly contained in the first component when the tRNA is charged. For verification, PCA was performed on the last 5 ns of the two 20 ns simulations of free $tRNA^{Glu}$ and Glu-$tRNA^{Glu}$. Figure 3.12 confirms the trend with the uncharged tRNA motion split among the first three components while the charged tRNA motion is almost completely contained within the first component. This suggests that charging causes a fundamental change in the motion of tRNA, which may be a necessary characteristic for dissociation.

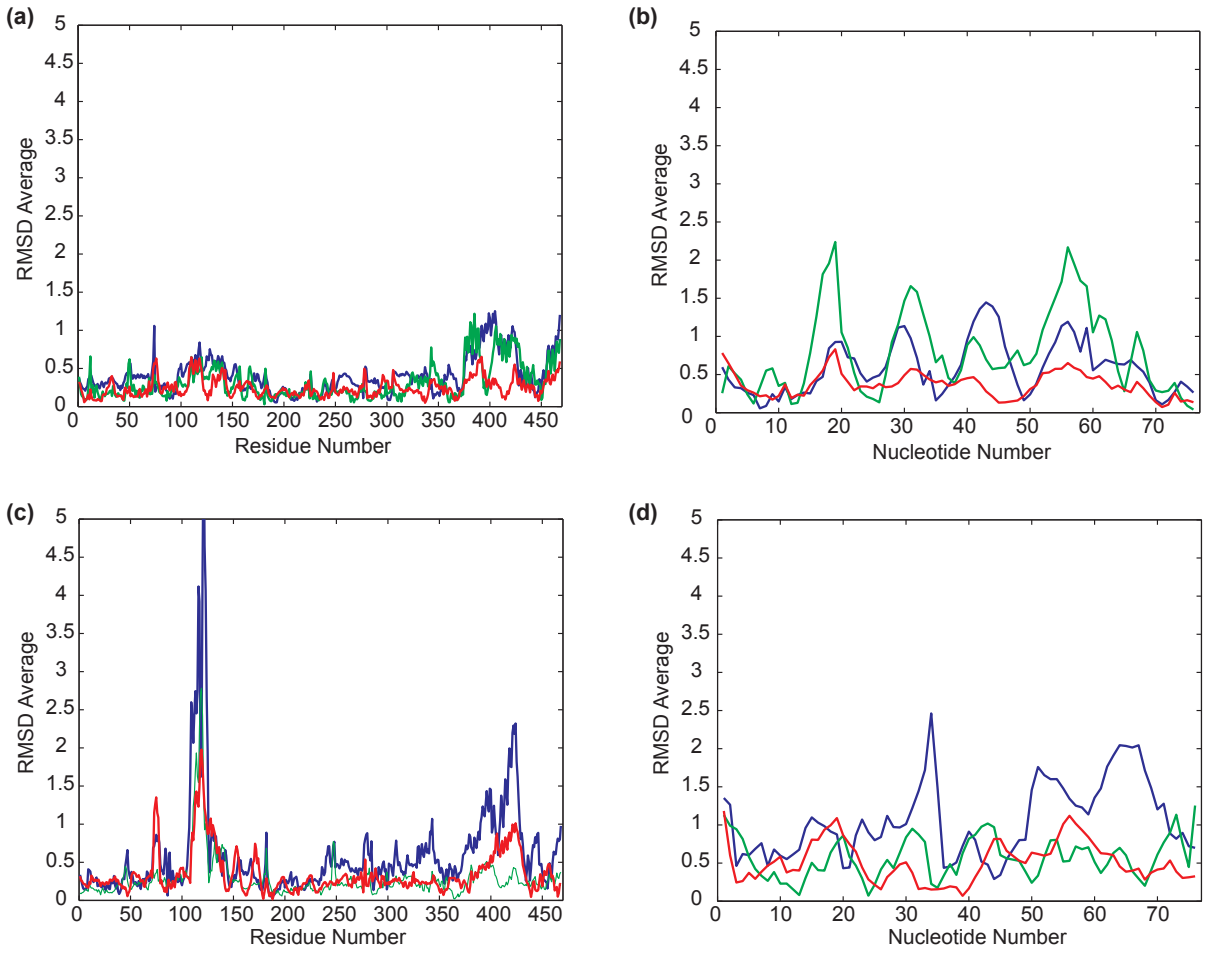


Figure 3.11: Comparison of the average RMSD for each residue or nucleotide over the last 5 ns of the Pre-transfer state GluRS (a) and $tRNA^{Glu}$ (b) and Post (no $AMP/GluNH_2$) state GluRS (c) and Glu-$tRNA^{Glu}$ (d) trajectories. Motion in the first, second, and third principal components is shown in blue, green and red respectively.

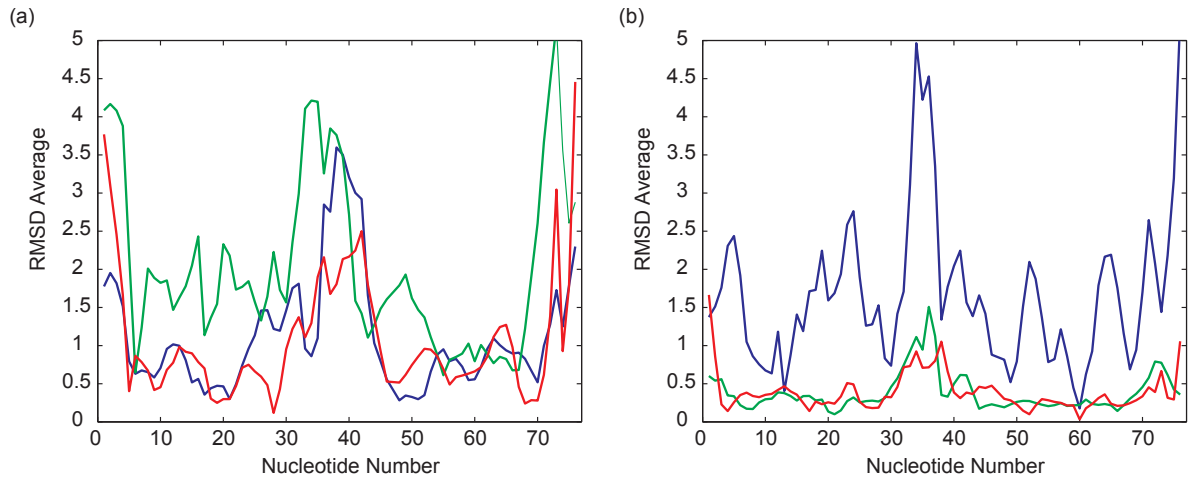


Figure 3.12: Comparison of the average RMSD for each nucleotide over the last 5 ns of the free $tRNA^{Glu}$ (a) and Glu-$tRNA^{Glu}$ (b) transformed system trajectories. Motion in the first, second, and third principal components is shown in blue, green and red respectively.

3.4.4 Communication decreases in interaction network as tRNA dissociates

Analysis of the dynamical network of interactions within the aaRS·tRNA complexes was used to further identify signs of dissociation during the 20 ns simulations (relatively short compared to the biological timescale of tRNA release) [131]. This type of network analysis uses local coupled motions between pairs of residues to track allosteric signaling in biomolecules. A node represents either a nucleotide, an amino acid, or the small molecule substrate/product present in the protein·RNA complex, and edges connect nodes that are in contact for a majority of the last 5 ns of the trajectory (see Methods). The cross-correlations calculated from the atomic fluctuations during the same time window are used to weight edges such that as the correlation reduces, the edge distance lengthens. Greater edge distances indicate lower communication between connected pairs of nodes since the corresponding residues are moving more independently of one another. Signaling pathways are dominant routes for communication between important interfacial contacts and the active site, and they are found by tracing the most optimal path through weighted edges in the network. Shorter optimal path lengths have greater correlation along the path. The sum of all path lengths from the interfacial interactions to the active site are therefore

used to quantitatively compare changes in correlation during tRNA release. Reduction of correlation with time during the release of the charged tRNA from the protein may also be observable experimentally through mutagenesis studies.

The shortest paths shown in Table 3.3 are calculated between A76 (site of aminoacylation) and identity elements in the tRNA. Identity elements are specificity dependent nucleotides that have been experimentally shown to affect the efficiency of aminoacylation by the cognate aaRS [208, 65, 382]. The $tRNA^{Glu}$ identity elements are divided into three groups according to their position on the nucleic acid: acceptor stem (G1·C72, G2·U71, and C4·G69), core (U11·A24, C12·G23··C9, and U13·G22··A46), and anticodon arm (C34, U35, C36, and A37). Path distances are summed over all identity elements in a group. The path length sum is comparable for the Pre-transfer and the Post (H-AMP) states, indicating that correlations in the post-transfer state immediately following the transition state are nearly the same as in the pre-transfer state. In all subsequent post-transfer states, correlations reduce along the signaling pathways, demonstrating that interfacial contacts are weakening. In the Post (AMP) and Post (no AMP) states, the signaling pathways have comparable distances with the largest change occurring at the acceptor stem, where the tRNA is starting to undock. No significant change is observed in the core and the anticodon arm, suggesting that the acceptor stem is released first, as hypothesized from the TrpRS·Trp-$tRNA^{Trp}$ crystal structure [315]. The Post (no AMP/$GluNH_2$) system has the largest total distance, indicating that this state is the closest to dissociation. Acceptor stem distances show the largest loss because many of the interactions between these nucleotides and GluRS have broken.

Source	Pre-transfer	Post (H-AMP)	Post (AMP)	Post (no AMP)	Post (no AMP/ $GluNH_2$)
G1	4.80	-	7.67	7.05	15.42
C72	1.35	2.24	3.59	5.90	13.86
G2	2.50	5.51	5.92	4.51	12.26
U71	1.94	2.48	2.98	2.63	12.97
C4	2.25	3.69	3.97	3.03	9.85
G69	3.30	2.87	2.78	2.78	8.26
AS Subtotal	16.14	16.79	26.91	25.90	72.62
U11	4.55	3.09	3.16	2.88	9.15
A24	3.79	3.35	3.66	3.36	10.25
C12	3.30	2.65	2.63	2.59	8.61
G23	4.07	3.47	4.46	3.68	9.62
C9	4.29	3.06	3.15	3.15	9.42
U13	3.32	2.80	3.19	2.75	8.84
G22	4.22	4.22	3.99	4.11	10.06
A46	4.80	3.68	3.96	3.85	9.81
Core Subtotal	32.34	26.32	28.20	26.37	75.76
C34	5.44	5.25	5.19	6.09	12.34
U35	4.71	4.33	4.18	5.72	11.18
C36	4.25	4.00	6.40	5.93	10.69
A37	3.96	3.60	5.16	4.71	10.58
Anticodon Arm Subtotal	18.36	17.18	20.93	22.45	44.79
Total	66.84	60.29	76.04	74.72	193.17

Table 3.3: Comparison of optimal path distances between different system states. For each state, the shortest distances between A76 and identity elements are listed; the identity elements are divided into groups based on their structural position. The distances within each group are summed with the final total at the bottom. Shorter distances indicate stronger signaling. Dashes indicate a nucleotide has lost all contacts with its neighbors and therefore completely isolated.

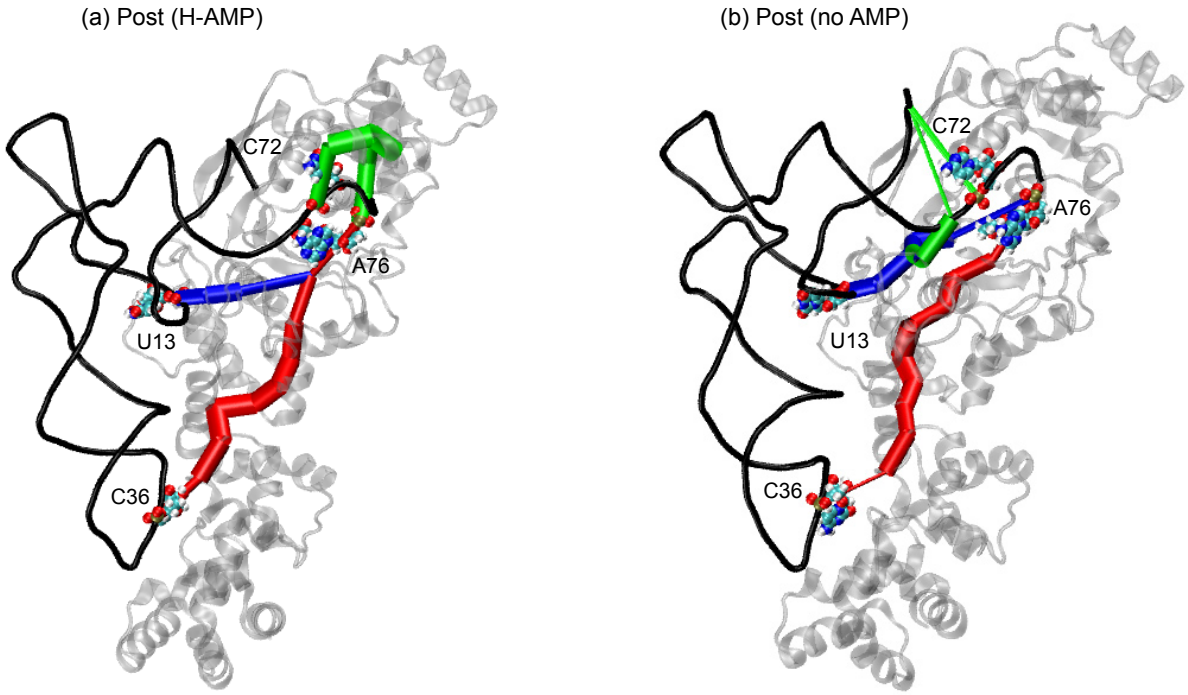


Figure 3.13: Comparison between the dynamical networks for the Post (H-AMP) and Post (no AMP) states. The correlations from the last 5 ns of the 20 ns simulation of the Post (H-AMP) and Post (no AMP) states were used to calculate the dynamical network (see Methods). The optimal paths are shown between A76 and identity elements C72 (green), U13 (blue), and C36 (red). Thicker edges indicate higher correlation between two nodes. The green and red paths are longer and thinner in (b) than in (a) revealing that the Post (no AMP) state complex has lost correlation between the charging amino acid on A76 and nucleotides C36 and C72 (see Table 3.3). The blue path, however, is only slightly thinner in (b) indicating that signaling is comparable for U13 between the two states.

In Figure 3.13 and 3.14, the shortest paths between A76 and the identity elements C72, U13, and C36 are compared for the Post (H-AMP) and Post (no AMP) states. Paths are traced along the nodes, and the thickness of each edge is scaled by the correlation weights (thicker edges show greater correlation). As seen in the representative optimal path between C72 and A76, in the Post (H-AMP) state, the acceptor stem communicates with the active site through the CP1 insertion of the protein. In the Post (no AMP) state, the contacts between the CP1 insertion and the acceptor stem have broken as the CCA hairpin unbends and begins to unbind. Instead, the optimal path length increases as it passes through the tRNA before merging with the shortest path from U13 to A76 that passes

through the catalytic domain of the protein. The shortest paths from U13 and C36 merge as they pass through the AMP before reaching A76 in the Post (H-AMP) state. The absence of the AMP causes a shift in these shortest paths in the Post (no AMP) state even though the anticodon loop and the D arm do not undock significantly. As discussed in the next section, the dissociation of the charged tRNA involves a conformational twist that brings the charged glutamate and CCA hairpin out of the active site while the D and anticodon stems are driven slightly further into the GluRS. In the Post (no AMP/$GluNH_2$) state, the partially undocked CCA hairpin and charging amino acid have lost nearly all communication with nonneighboring residues/nucleotides. The identity elements in the anticodon and D arms remain in contact with GluRS residues, but because the number of contacts between the partially undocked CCA hairpin and the protein is significantly reduced, longer paths are required to pass communication to A76.

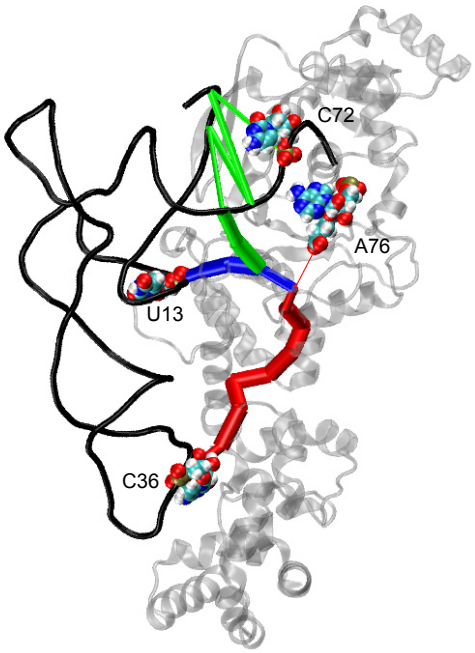


Figure 3.14: Dynamical network of the Post (no AMP/$GluNH_2$) state. The shortest or optimal paths are shown between A76 and nucleotides C72 (green), U13 (blue), and C36 (red). Thicker edges indicate greater correlation between two nodes and the thickness of the path between Lys241 and A76 has been increased for visibility.

3.4.5 Energetics of GluRS·tRNA^{Glu} binding interface

GluRS specifically recognizes and binds tRNA^{Glu} through contacts spanning the acceptor stem to the anticodon, and these interactions must break during tRNA dissociation. The residues and nucleotides making significant energetic contributions at the interface in the Pre-transfer state shown in Figure 3.15 were identified by averaging the electrostatic and van der Waals interaction energies over the last 5 ns of each simulation. These energies were calculated with a cutoff of 21 Å to fully include bound tRNA helices while limiting the analysis to the GluRS binding interface. The evolutionary importance of these interface residues is shown by masking the contact energy values by percent sequence identity across the bacterial GluRS evolutionary profile.

Because tRNA is highly electronegative, positively charged residues on GluRS generally form attractive contacts while negatively charged residues repel the tRNA. Figure 3.15a shows twice as many significantly attractive interactions as repulsive for the Pre-transfer complex, and all major non-bonded interactions involve charged amino acids near the highly negative tRNA backbone phosphates. The attractive interactions are generally stronger than the repulsive because positively charged residues can form direct contacts with the tRNA while repulsive forces keep the negatively charged residues further away. Most of the strongly attractive residues are close in sequence to repulsive residues, energetically balancing the interactions such that the tRNA is not bound too tightly to dissociate from GluRS. The strongly attractive residues at the tRNA interface are also more highly conserved than the repulsive residues (59% conservation vs. 43% with background GluRS conservation at 41%). Almost half of these attractive amino acids have greater than 75% sequence identity whereas only one of the repulsive residues is similarly well conserved. Six of the strongly interacting residues contact identity elements on the tRNA through salt bridges to backbone phosphates or hydrogen bonds to polar groups on the ribose or base (see Tables 3.4 and 3.5).

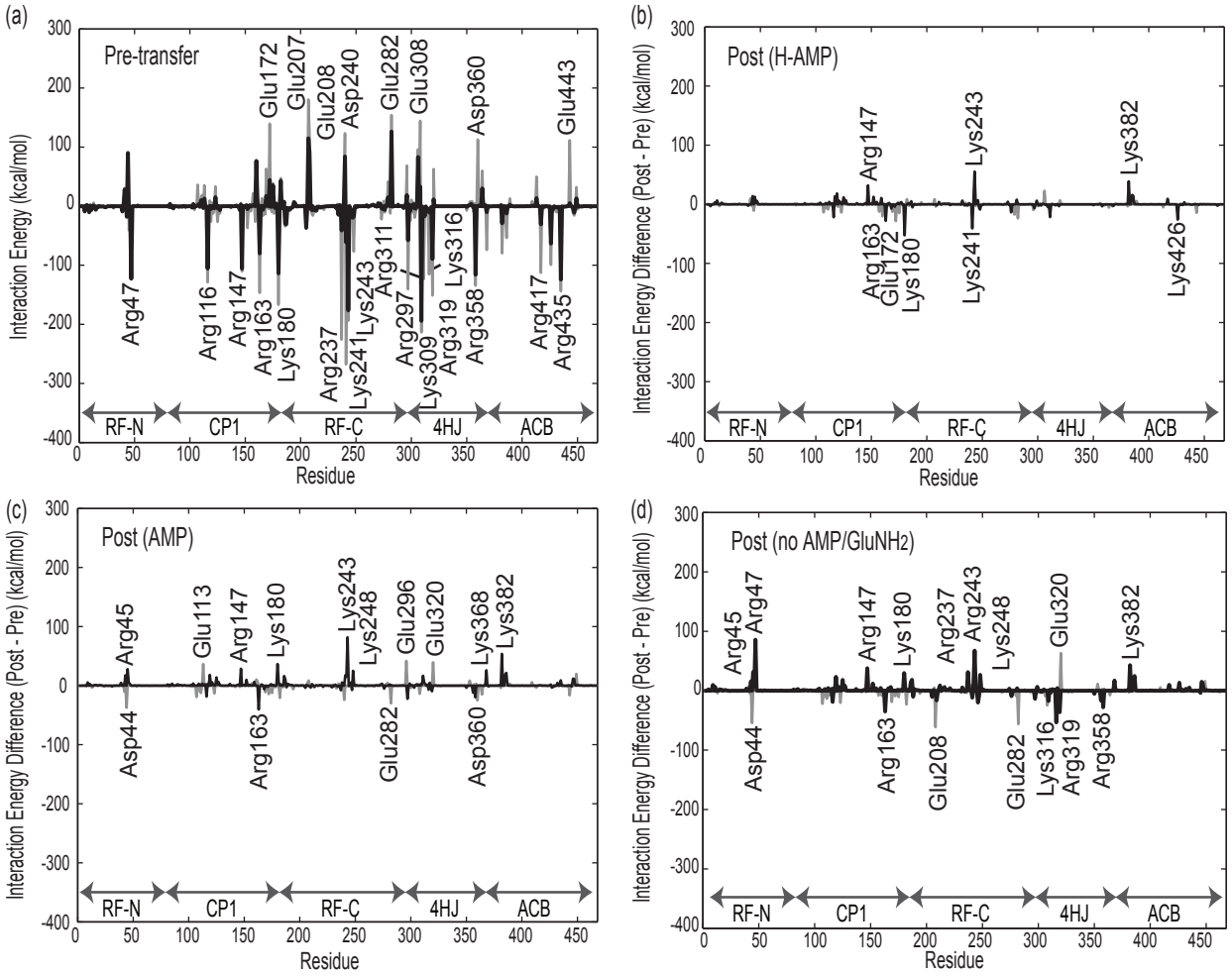


Figure 3.15: Mean nonbonded interaction energies between $tRNA^{Glu}$ and GluRS. The energies are averaged over the last 5 ns of the 20 ns simulation of the Pre-transfer state (a). Gray peaks show the full energetic interaction while black peaks show the energy scaled by percent sequence identity (see Methods). Labeled peaks have absolute value greater than 100 kcal/mol. The mean non-bonded interaction energy difference for GluRS to $tRNA^{Glu}$ when compared to (a) is shown for the (b) Post (H-AMP), (c) Post (AMP), and (d) Post (no $AMP/GluNH_2$) states (atoms in the charging glutamate were not included). Residues that make attractive or repulsive interactions in the Pre-transfer state are shown in black or gray, respectively. Black peaks greater than zero represent residues moving further away from $tRNA^{Glu}$ while positive gray peaks correspond to residues moving closer in. The opposite is true for negative peaks. These values are averaged over the last 5 ns of the 20 ns trajectories.

Identity Element	Residues forming direct contacts	Residues within 5 Å
Pre-transfer		
G1	none	none
C72	none	Glu172
G2	none	Glu172
U71	Lys243 (bb/sc)	Glu208
C4	none	none
G69	Arg237 (rib/sc) Lys243 (rib/bb)	Glu207 Lys241
U11	none	Lys241 Lys309
A24	Glu282 (rib/sc) Lys309 (rib/sc)	none
C12	Glu282 (rib/sc) Lys309 (rib/sc)	none
G23	Glu282 (base/sc)	Lys309
C9	none	none
U13	none	Arg297
G22	none	none
A46	none	none
C34	Arg417 (bb/sc) Arg435 (base/sc)	none
U35	none	Arg358 Arg435 Glu443
C36	Arg358 (base/sc)	Glu443
A37	none	Arg311 Arg319 Arg358 Glu443

Table 3.4: Important GluRS residues interacting with $tRNA^{Glu}$ identity elements. The identity elements are grouped similarly to Table 3.3. The residues were selected from those shown in Figure 3.15 to make highly energetic contacts with the tRNA. The parts of the nucleotide and residue forming the direct contacts are denoted in parenthesis with the following abbreviations: bb = phosphate backbone, rib = ribose, and sc = sidechain. The residues within 5 Å were selected from the crystal structure and end of the Post (no AMP/$GluNH_2$) simulation (20 ns).

Identity Element	Residues forming direct contacts	Residues within 5 Å
Post (no AMP/GluNH$_2$)		
G1	none	none
C72	Arg116 (rib/sc)	none
G2	Glu172 (rib/sc)	none
U71	none	none
C4	none	none
G69	Glu207 (base/sc) Arg237 (rib/sc)	Lys241 Lys243
U11	Lys241 (bb/sc)	Lys309
A24	Glu282 (rib/sc) Lys309 (rib/sc)	none
C12	none	Lys309
G23	none	Lys309
C9	none	none
U13	none	none
G22	none	none
A46	none	none
C34	Arg417 (bb/sc) Arg435 (base/sc)	none
U35	none	Arg358 Arg435 Glu443
C36	Arg358 (base/sc)	Glu443
A37	none	Arg319 Arg358

Table 3.5: Important GluRS residues interacting with tRNA^{Glu} identity elements (continued). The identity elements are grouped similarly to Figure 3.3. The residues were selected from those shown in Figure 3.15 to make highly energetic contacts with the tRNA. The parts of the nucleotide and residue forming the direct contacts are denoted in parenthesis with the following abbreviations (bb = phosphate backbone, rib = ribose, sc = sidechain). The residues within 5 Å were selected from the crystal structure and end of the Post (no AMP/GluNH$_2$) simulation (20 ns).

The strongest interactions occur in the second half of the Rossman fold, particularly near the KMSK motif (residues 243-246), which binds the 3′ side of the acceptor stem, and in

the 4HJ near the D stem and the 5′ end of the anticodon stem. The stronger attractive interactions limit the rate at which dissociation can occur. Finally, the anticodon and its neighboring nucleotides are bound by Arg358 (86% with 14% lysine), Arg417, and Arg435 (86%) primarily through interactions with functional groups on the nucleobases while being counterbalanced by repulsion from nearby Asp360 and Glu443. Arg358 (Lys358) is notable because the presence of this positively charged residue is used by bacterial D-GluRS to discriminate between $tRNA^{Glu}$ and $tRNA^{Gln}$ [327]. While the charged CCA end leaves the GluRS catalytic site, the tRNA elbow region undergoes a rocking motion that moves the acceptor stem further away from GluRS and the D stem closer.

Nonbonded interactions to charging glutamate A similar analysis was performed in the active site between GluRS and the charging glutamate moiety attached to the tRNA (see Figure 3.16). In the Pre-transfer state, the three most attractive contacts to the charging glutamate are through salt bridges from the glutamate sidechain to Arg5 and Arg205 and from its α–ammonium group to the conserved Glu41 (see Figure 3.8 for structural reference). In the Post (H-AMP) state, the interactions are almost identical, but the contact to Arg5 has been lost in the Post (AMP) state. Although Arg5 and Arg205 interact strongly with the glutamate during transfer, both of their sidechains are extended and accessible to water. Water molecules intervene between Arg5 and the charging glutamate, breaking the corresponding salt bridge. This state is likely to be in the process of dissociating, as seen in the previous section, but with the glutamate remaining bound to two of its three primary contacts, it does not progress at the same rate as the Post (no $AMP/GluNH_2$) state. The glutamate contacts differ significantly in the partially undocked Post (no $AMP/GluNH_2$) state. As the charging glutamate leaves the active site, water molecules come between it and the active site residues, and all three original salt bridges are broken. As the CCA end pivots around the phosphoester linkage between nucleotides 70 and 71, the glutamate forms new contacts to Arg47 and positively charged residues Lys243, Lys246, and Arg247 in the KMSK loop. During this simulation, the sidechain of the charging glutamate makes transient salt bridges with all four of these residues. The

longest-lived salt bridges are made between the glutamate and Lys246.

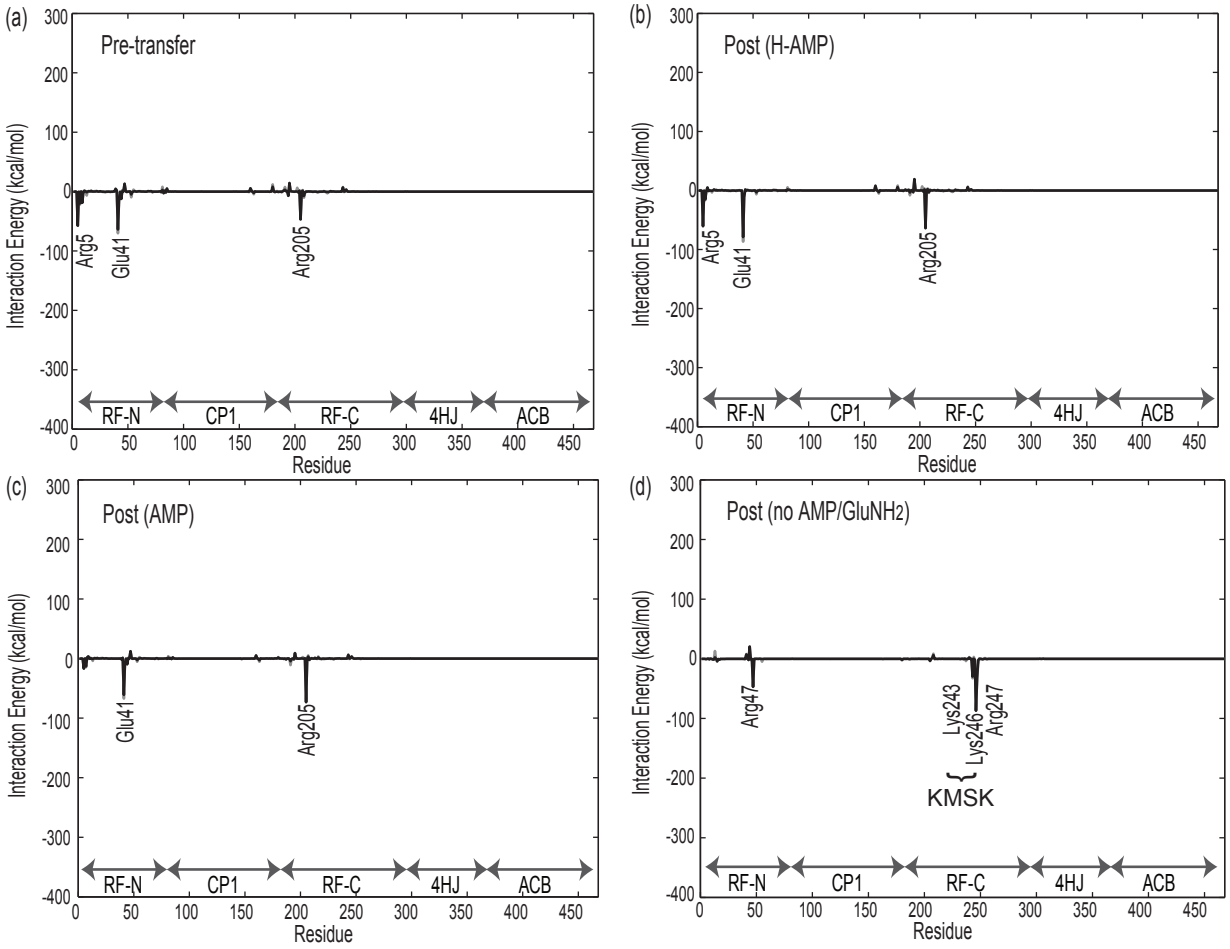


Figure 3.16: Mean nonbonded interaction energies between the charging amino acid (glutamate) and GluRS in the Pre-transfer, Post (H-AMP), Post (AMP), and Post (no AMP/$GluNH_2$) states. Labeled residues have interaction energies greater than 25 kcal/mol. These values are averaged over the last 5 ns of the 20 ns trajectories.

3.4.6 Free energies of binding predict tRNA dissociation irrespective of AMP

The technique of Molecular Mechanics–Poisson-Boltzmann Surface Area (MM-PBSA) is used to calculate both the average free energy of complex formation and binding of ligands [366, 367], $\langle \Delta G_{\text{binding}} \rangle = \langle G_{AB} \rangle - (\langle G_A \rangle + \langle G_B \rangle)$ where AB denotes a macromolecular complex formed from components A and B. This can be reformulated with

the free energy difference expressed as a sum of terms calculated from MD trajectories: $\langle \Delta G_{binding} \rangle = \langle \Delta E_{VdW} + \Delta E_{elec} + \Delta G_{polar} + \Delta G_{nonpolar} \rangle - T\langle \Delta S \rangle$. The $\langle \cdot \rangle$ indicate time averages over the last 5 ns of each 20 ns trajectory. The first two terms are used to measure the van der Waals and electrostatic interaction energies between the two molecules and are derived from molecular mechanics. G_{polar}, the polar solvation energy, is the energy required to move the system from a dielectric of $\epsilon_{in} = 1.00$ to $\epsilon_{out} = 78.54$. The polar solvation energy is efficiently calculated using a Poisson-Boltzmann implicit continuum solvent model [372]. The nonpolar solvation energy, $G_{nonpolar}$, is the energy required to create a cavity in the solvent for a given system and is proportional to the solvent accessible surface area. The last term refers to the change in entropy on complex formation and was calculated from the MD simulations using Schlitter's formula [373]. Since the free energies are calculated using the complex and system components from the same trajectory, free energies associated with any conformational changes in the isolated components are neglected in this analysis. Also, as MM-PBSA is not generally successful at reproducing absolute binding free energy values accurately, the focus here is on the trends in binding free energies between different system states.

The network analysis indicated that the loss of correlation is comparable in the Post (AMP) or Post (no AMP) states. To determine which is energetically more feasible to occur first - the dissociation of AMP (GluRS·Glu-tRNA^{Glu}·AMP $\rightleftharpoons$ GluRS·Glu-tRNA^{Glu} + AMP) or charged tRNA (GluRS·Glu-tRNA^{Glu}·AMP $\rightleftharpoons$ GluRS·AMP + Glu-tRNA^{Glu}) - two sets of calculations were performed and reported in Tables 3.7 and 3.8, respectively. In the first, the free energy of binding was estimated for the small molecule substrate/products (AMP or H-AMP) in the Pre-transfer, Post (H-AMP), and Post (AMP) states, and in the second, the free energy of binding was estimated for charged tRNA to GluRS in the presence and absence of the appropriate small molecule. The experimentally measured free energy of binding of tRNA to homologous class I aaRS enzymes in the absence of any cofactors is -8 to -9 kcal/mol [383, 76]. The interior dielectric was chosen to make the Pre-transfer $\langle \Delta G^{tRNA}_{binding} \rangle$ have a comparable value (see Table 3.6 for the detailed calculation).

Pre-transfer	Complex	GluRS/tRNA	Adel	Δ	Complex	GluRS/Adel	tRNA	Δ
$\langle E_{VdW} \rangle$	-2041.96	-2013.15	12.54	-41.34	-2041.96	-1996.16	198.69	-244.49
	(1.86)	(1.83)	(0.12)	(0.20)	(1.86)	(1.45)	(1.20)	(0.34)
$\langle E_{elec} \rangle$	-235075.29	-233086.62	-1984.77	-3.90	4560.52	-16559.16	21861.89	-742.21
	(4.09)	(4.05)	(0.38)	(0.38)	(9.63)	(4.70)	(7.83)	(6.42)
$\langle E_{coulomb}+G_{polar} \rangle$	-41465.21	-40804.52	-173.61	-487.08	-235075.29	-140446.75	-94857.11	228.57
	(9.19)	(8.98)	(0.17)	(0.50)	(4.09)	(2.91)	(2.69)	(0.74)
$\langle G_{non-polar} \rangle$	166.35	167.54	4.62	-5.81	166.35	121.71	73.71	-29.07
	(0.05)	(0.05)	(0.00)	(0.00)	(0.05)	(0.04)	(0.03)	(0.02)
-TΔS	-1242.30	-1238.00	-16.20	11.90	-1242.30	-968.70	-309.80	36.20
$\langle G_{binding} \rangle$	-234347.91	-232347.90	-1960.86	-39.15	-234347.91	-140590.98	-93748.14	-8.79
	(4.27)	(4.25)	(0.38)	(0.29)	(4.27)	(3.24)	(2.71)	(0.65)

Table 3.6: Detailed MM-PBSA binding free energy values for the Pre-transfer state. The first set of columns show the values resulting in $\langle \Delta G_{adenylate} \rangle$ of -39.15 kcal/mol. The second set correspond to $\langle \Delta G_{tRNA} \rangle$. The 95% confidence interval range for each quantity is ± the value shown in parentheses.

Small molecule substrate undocking The small molecule product binding free energies indicate that H-AMP/AMP can dissociate from the GluRS·Glu-tRNA^{Glu} complex after the reaction (See Table 3.7). The adenylate binds tightly to the GluRS·tRNA^{Glu} component while both H-AMP and AMP have a significantly reduced affinity for complex formation. This results from a change in the ΔG_{polar} + $\Delta E_{coulomb}$ due to the transfer of the glutamate moiety to the tRNA. The charging glutamate backbone forms a salt bridge with Glu41 while its sidechain carboxylate forms two salt bridges with Arg5 and Arg205, accounting for the strong binding of the adenylate to active site residues. In comparison, the AMP moiety forms relatively weaker hydrogen bonds with protein residues in the active site (see Figure 3.6). In addition, the transfer of glutamate to tRNA reduces the surface area (also leading to smaller $\Delta G_{nonpolar}$) as well as the charge distribution on the interaction surface of the complex. Conformational and charge differences on the small molecule product account for the AMP being more likely to exit the active site than H-AMP. The 5 kcal/mol difference in the ΔE_{vdW} between the Post (H-AMP) and Post (AMP) states is due to the AMP shifting in the pocket and making contact with fewer residues, particularly His15. The difference in charge from H-AMP to AMP increases the contribution of the

ΔG_{polar} term. The proton transfer causes the charge of H-AMP/AMP to change from -1 to -2 while the charge of the GluRS:Glu-tRNA^{Glu} complex remains -71. This results in a larger coulombic repulsion after proton transfer, compensating for the change in ΔG_{polar}.

Small Molecule Substrate Free Energies of Binding

	Pre-transfer	Post (H-AMP)	Post (AMP)
$\langle \Delta E_{VdW} \rangle$	-41.34	-41.79	-36.77
	(0.20)	(0.13)	(0.13)
$\langle \Delta E_{coulomb} + \Delta G_{polar} \rangle$	-3.90	39.47	38.47
	(0.38)	(0.32)	(0.29)
$\langle \Delta G_{nonpolar} \rangle$	-5.81	-4.85	-4.72
	(0.00)	(0.00)	(0.00)
$-T\Delta S$	11.70	8.40	8.40
$\langle \Delta G_{binding} \rangle$	-39.15	1.23	5.38
	(0.29)	(0.31)	(0.27)

Table 3.7: MM-PBSA free energy estimates in kcal/mol for the adenylate/(H-)AMP substrate binding GluRS·tRNA^{Glu}. The 95 % confidence interval range is ± the number shown below in parentheses. Standard deviations for the $\Delta G_{binding}$ were all 6-7 kcal/mol.

TRNA dissociation The overall trends in binding energies between tRNA and GluRS in the various states seen in Table 3.8 are similar to those exhibited by the reduction in correlation in the network analysis. The Pre-transfer and Post (H-AMP) states have strongly negative $\langle \Delta G^{tRNA}_{binding} \rangle$ values that indicate tighter binding than in the other post-transfer states. The main contribution to the decrease of binding free energy comes from the change in the van der Waals interaction, indicating a decrease in the contacts between tRNA and the GluRS. Notably, this value is least attractive in the Post (AMP) and Post (no AMP/GluNH$_2$) states, both of which were identified as having the weakest interface contacts by both the network and local energetics analyses. With the exception of the Post (H-AMP) state, these results are in contradiction to the experiments on CysRS [76] that report a larger binding affinity to the charged tRNA^{Cys} than to the uncharged tRNA. The results indicate that charged tRNA can dissociate from GluRS in the presence or absence of AMP. Given that AMP can also undock on transfer of the charging amino acid to tRNA, the free energy calculations suggest that undocking of products from GluRS can proceed along competitive thermodynamic pathways, which could lead to kinetic partitioning. The only 20 ns simulation in which significant undocking of the tRNA

acceptor stem from GluRS was actually observed was in the Post (no AMP/$GluNH_2$) state.

tRNA Free Energies of Binding

	Pre-transfer	Post (H-AMP)	Post (AMP)	Post (no AMP)	Post (no AMP/ $GluNH_2$)
$\langle \Delta E_{VdW} \rangle$	-244.49	-234.03	-209.52	-215.70	-206.53
	(0.34)	(0.45)	(0.31)	(0.50)	(0.35)
$\langle \Delta E_{coulomb} + \Delta G_{polar} \rangle$	228.57	213.26	214.16	213.06	204.70
	(0.74)	(1.07)	(0.95)	(1.09)	(0.79)
$\langle \Delta G_{nonpolar} \rangle$	-29.07	-30.74	-28.20	-28.05	-26.34
	(0.02)	(0.04)	(0.03)	(0.03)	(0.03)
$-T\Delta S$	36.20	37.10	38.60	37.20	37.50
$\langle \Delta G_{binding} \rangle$	-8.79	-14.41	15.04	6.51	9.33
	(0.65)	(0.91)	(0.84)	(0.85)	(0.65)

Table 3.8: MM-PBSA free energy differences in kcal/mol for $tRNA^{Glu}$ binding to GluRS (with small molecule substrate/product). The 95 % confidence interval is ± the number shown below in parentheses. Standard deviations for the $\Delta G_{binding}$ were all 14-21 kcal/mol.

	Post (AMP/H15h)	Post (AMP/E41COOH/$GluNH_2$)
$\langle \Delta E_{VdW} \rangle$	-237.11	-244.46
	(0.45)	(0.37)
$\langle \Delta E_{coulomb} + \Delta G_{polar} \rangle$	219.42	233.71
	(1.00)	(1.07)
$\langle \Delta G_{nonpolar} \rangle$	-30.67	-31.14
	(0.02)	(0.03)
$-T\Delta S$	37.10	37.80
$\langle \Delta G_{binding} \rangle$	-11.26	-4.10
	(0.77)	(1.03)

Table 3.9: MM-PBSA free energy differences for tRNA binding in kcal/mol for alternative system states containing AMP. The 95 % confidence interval for each value is ± the number shown below in parentheses. Standard deviations for the $\langle \Delta G_{binding} \rangle$ were 17-24 kcal/mol.

Free energies of binding as a function of time Free energies of binding were calculated from the last 5 ns of the 20 ns simulations to capture the short time binding that is characteristic of each system state. The total $\Delta G(t)$ over time is shown in Figure 3.17 for the first 20 ns and the last 5 ns of the 80 ns Post (no AMP/$GluNH_2$) simulation. The average free energy over the first 5 ns of this simulation is roughly equivalent to the final 5 ns of the Pre-transfer state, indicating that the Post (no AMP/$GluNH_2$) state is still inhabiting the ensemble of states similar to the Pre-transfer state. However, the free energy of binding rapidly rises as the system equilibrates the modeled active site changes, peaking at 17.69 kcal/mol between 10 and 15 ns. During this

time, the charging amino acid exits the active site and the CCA hairpin begins to undock. The free energy becomes less repulsive during the next 5 ns as the charging amino acid makes surface interactions with the KMSK loop, stabilizing the CCA hairpin. During the final 5 ns, each of the individual components has moved closer to zero, as expected. As the interface opens, ΔE_{vdW} and $\Delta G_{nonpolar}$ weaken because the surface of each molecule becomes more exposed to solvent. There may be increasing error in the electrostatic components related to the treatment of the bulk dielectric as water enters the interface and creates a cleft between the two macromolecules, but the generally positive free energies indicate that Glu-tRNA^{Glu} is dissociating from GluRS.

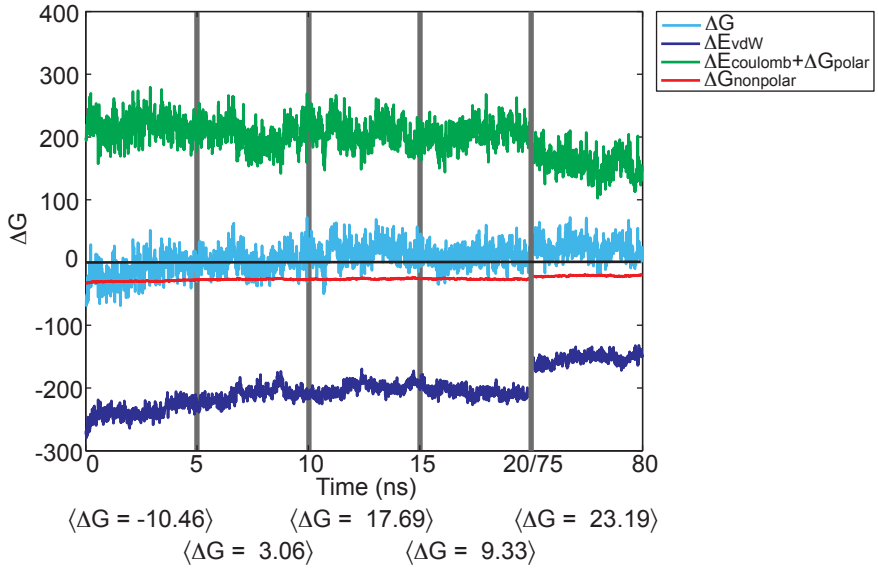


Figure 3.17: $\Delta G(t)$ averaged over 5 ns windows from the 80 ns Post (no AMP/GluNH$_2$) simulation for charged tRNA dissociation from GluRS. The components of the total $\Delta G(t)$ sampled every 40 ps are given over the first 20 ns and the last 5 ns of the 80 ns simulation. The entropy is averaged over 5 ns windows and included in $\Delta G(t)$.

3.4.7 Exit strategies for dissociation of charged tRNA

Figure 3.18 depicts several of the representative states as a series of events leading to tRNA dissociation. The first event is aminoacylation, the transition between the pre- and post-transfer states. The pre-transfer state (b) can be well approximated from the crystal

structure containing the analog, and the Post (H-AMP) state (c) previously postulated to exist [317] following transfer of the amino acid moiety. Network analysis and free energy calculations indicate that the Post (H-AMP) state complex is stable and has dynamical and energetic properties similar to the Pre-transfer state complex. Transfer of the 2′-hydroxyl proton to AMP causes significant conformational changes in the active site with the HIGH motif providing stability for the H-AMP. The formation of hydrogen bonds between H-AMP and both His15 and Ser9 reduces the pKa of H-AMP to 5.51, which sets up the next transition with the transfer of the proton to a general base.

While the HIGH motif would seem to be a good candidate for this transfer, pKa calculations show that His15 has a strong propensity for remaining neutral. The only other general base in the vicinity is water, which is modeled in the Post (AMP) state (d). From here, both the AMP and the charged tRNA can dissociate along competitive pathways. However, the CCA hairpin, which has been hypothesized to dissociate prior to the rest of the tRNA, remains strongly bound to the active site residues during the 20 ns simulation.

The Post (no AMP) state models the complex in the case where AMP has undocked before tRNA. The affinity of GluRS for the charged tRNA is similar to its affinity in the presence of AMP. The CCA hairpin becomes highly solvated by additional water molecules entering through the mobile KMSK loop and vacant AMP binding site. This helps reduce interactions between the charged CCA hairpin of the tRNA molecule and GluRS. However, the rocking motion that assists in weakening the contacts between the acceptor stem and the CP1 insertion also causes a conformational twist that strengthens the interactions between GluRS and the tRNA core. This results in a energetic trade-off which causes the Post (no AMP) and the Post (AMP) states to have comparable shortest path lengths.

The complete release of the charged CCA hairpin is prevented by contacts between the α-ammonium group of the charging amino acid and the conserved residue Glu41. On the transfer of the charging amino acid to the tRNA, the pKa of the α–ammonium group decreases to 7.23, while the pKa of Glu41 increases to 8.06, indicating that the proton

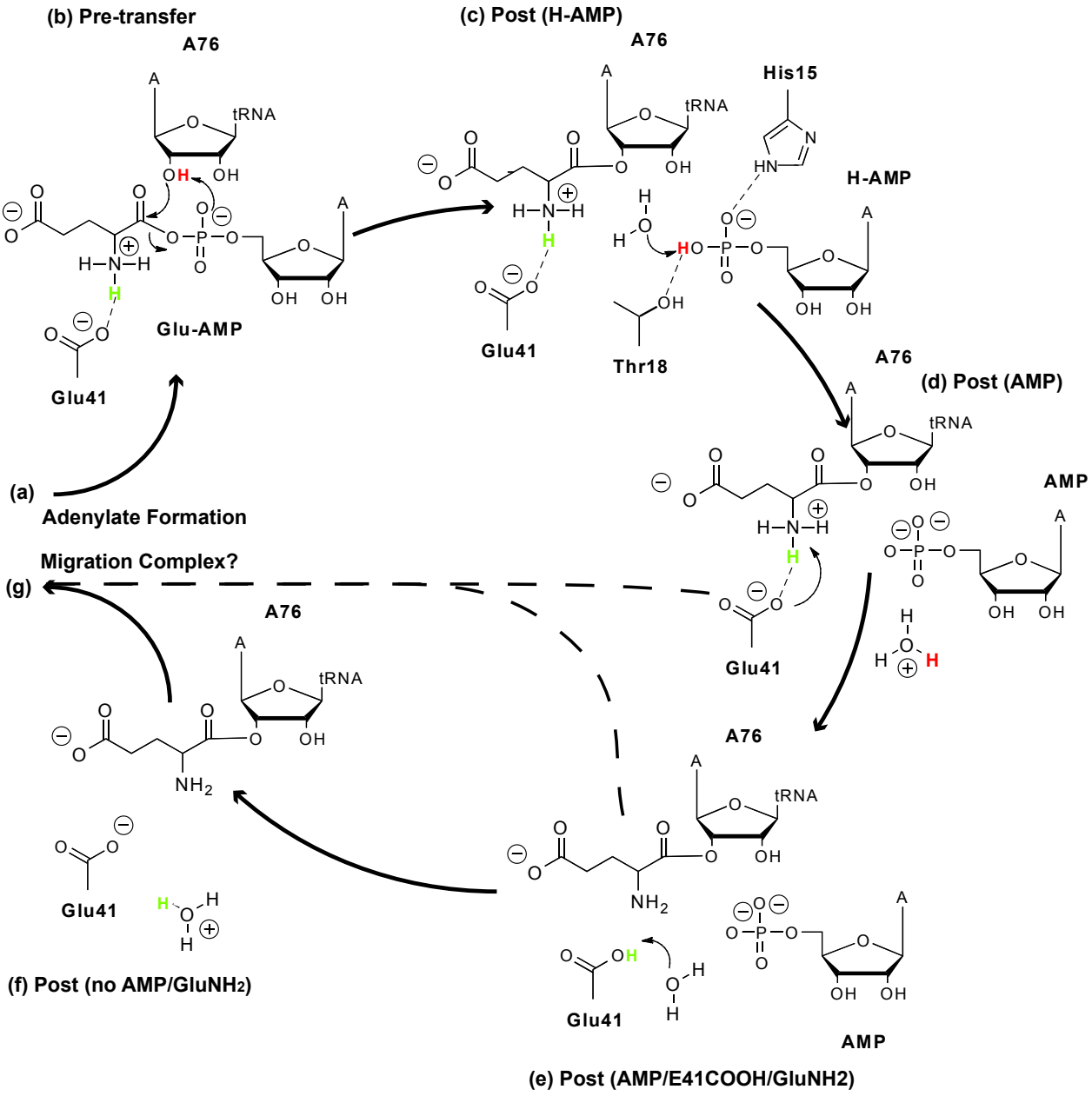


Figure 3.18: Proposed tRNA exit strategies. The various pathways can be constructed from the states considered in this study: each begins after the tRNA has associated with GluRS and the adenylate has been formed (a), and ends with the formation of a proposed migration complex with EF-Tu (g). Dotted arrows imply additional dissociation events that could occur.

could be transferred from the charging amino acid to Glu41, reducing the interactions between the two (Post (AMP/GluNH$_2$/E41COOH) state) (e). The removal of AMP creates a channel for water molecules to interact with Glu41COOH and solvate contacts between the charging amino acid and other active site residues. The proton can be transferred subsequently to the water molecules entering through the vacant AMP binding site. Removal of the proton allows the CCA hairpin and charged amino acid to exit the active site within 20 ns (Figure 3.16). The anticodon loop and core remain bound, while fluctuations in the acceptor stem increase.

The final state in Figure 3.18 is the proposed migration complex consisting of GluRS·tRNA^{Glu}·EF-Tu·GTP. The structural details of the docking of EF-Tu to the GluRS·tRNA^{Glu} complex are unknown, but recent studies have suggested that EF-Tu can bind the tRNA while it is still bound to the aaRS [78, 76]. The steric clashes reported upon docking of the crystal structures [76] are eliminated by selecting configurations from the MD simulations in which the CP1 insertion has rocked away from EF-Tu. Because the binding affinity of tRNA for EF-Tu is 300-fold higher than the aaRS [76], this might also serve as a powerful stimulant for tRNA release. We modeled the initial migration complex using equilibrated EF-Tu from *T. thermophilus* (from [63]) and the Post (no AMP) state complex after 20 ns. After 20 ns of equilibration, the dynamical network and community partitioning in the migration complex were calculated from the last 5 ns and compared to the network from the Post (no AMP) state simulation. Shown in Figure 3.19 are network communities comprised of the tRNA core and T arm. In the left panel showing the GluRS·tRNA^{Glu} in the absence of EF-Tu, four nucleotides in the tRNA core remain strongly correlated with the GluRS. However, in the presence of EF-Tu (right panel), the tRNA core gradually loses its correlations with the GluRS while the T arm becomes strongly correlated with the EF-Tu. While this simulated model of the migration complex is only one representative of the ensemble of partially docked states, it indicates that EF-Tu can have a strong stimulating effect on the release of the tRNA from the aaRS. Subsequent deprotonation of the α–ammonium group, which is still bound to conserved active site

residues, results in release of the CCA hairpin towards the binding site in EF-Tu (see Figure 3.20). Open questions remain with regards to the initial approach and binding of EF-Tu to the GluRS:tRNA complex. The structural details and timeline for EF-Tu binding could affect interactions between GluRS and tRNA both before and after aminoacylation.

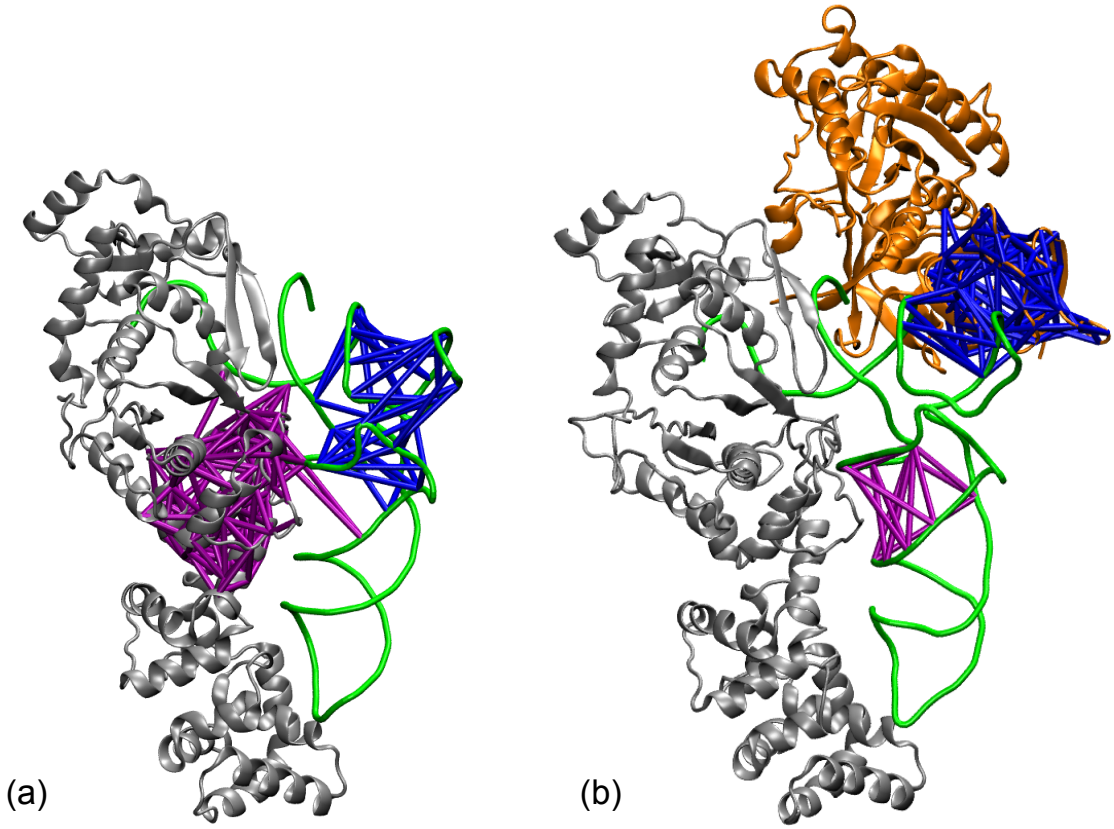


Figure 3.19: Community analysis of the (a) Post (no AMP) complex and (b) GluRS·tRNA^{Glu}·EF-Tu·GTP complex formed from the Post (no AMP) system. Two communities are displayed for each complex: one containing nucleotides in the the D stem (purple) and one containing the base of the T stem (blue). In the Post (no AMP) state, a community with four D stem nucleotides includes amino acids from the C-terminal half of the GluRS Rossman fold, and the T stem community contains the whole T arm. When EF-Tu binds the tRNA, the T stem community merges with the third domain of EF-Tu, and the D stem peels away from GluRS, resulting in a community containing only nucleotides. Dynamical networks were created from the final 5 ns of 20 ns trajectories. [*Figure courtesy of J. Eargle.*]

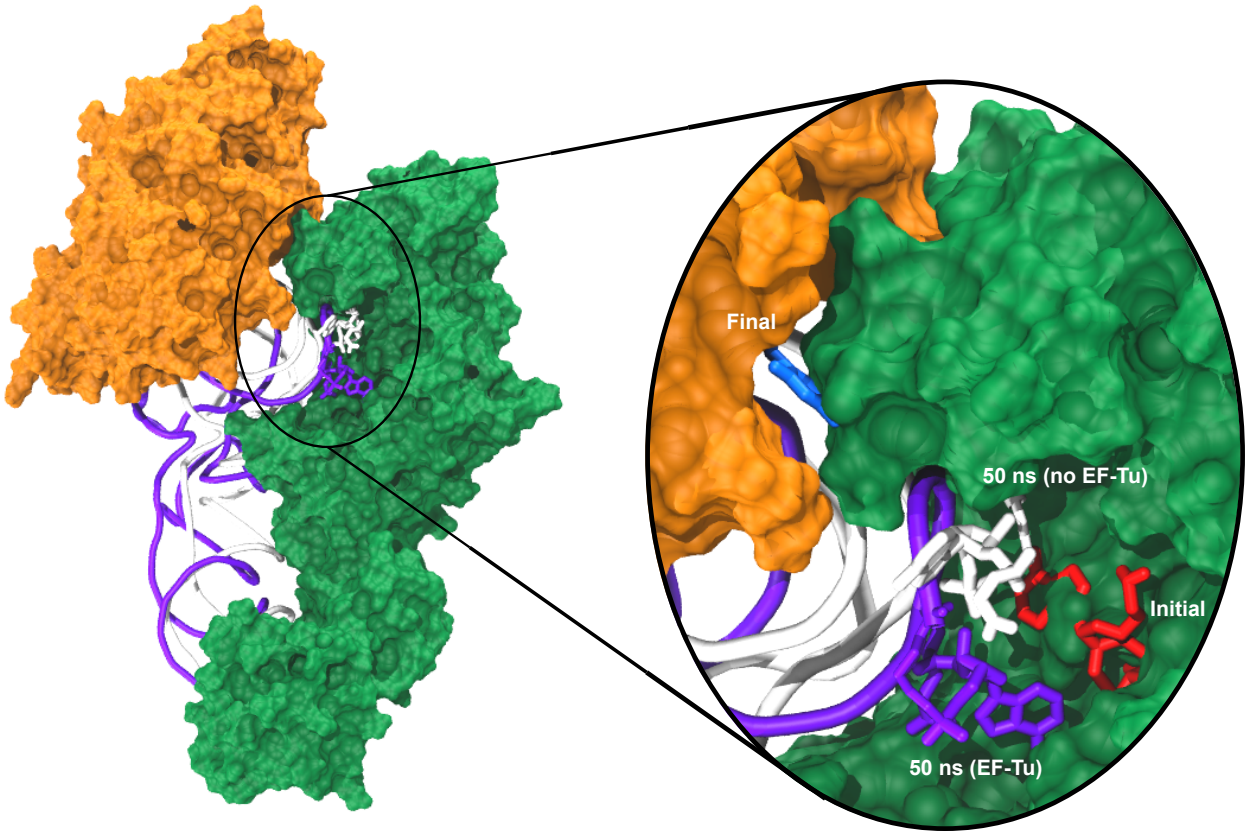


Figure 3.20: Motion of tRNA in the absence and presence of EF-Tu. The tRNA and charging amino acid are shown in the Post (no AMP/$GluNH_2$) and Post (no AMP/EF-Tu) states after 50 ns of equilibration. The Post (no AMP/EF-Tu) simulation started out with an α–ammonium group on the charging glutamate moiety. At 10 ns, this group was deprotonated, allowing the CCA end to undock from GluRS, and at 30 ns the α–amino group was reprotonated because it was surrounded by solvent. The equilibrated structures were aligned to the crystal structures of GluRS (1N78 [316]) and EF-Tu (1B23 [100]) by the protein backbones. The initial position of the charging amino acid in GluRS is shown in red, the position after 50 ns in the absence or presence of EF-Tu in white and purple respectively, and the position when docked to the EF-Tu (crystal structure) in blue (final).

3.5 Conclusion

Experiments reveal that the complete dissociation of the charged tRNA from the class I aaRSs takes place in the millisecond-second timescale and is stimulated by the presence of EF-Tu [76, 78], but our calculations indicate that there can be initial signs of tRNA release even at timescales of tens of nanoseconds. An important factor affecting the release of charged tRNA is the protonation state of residues in the active site of the aaRS. Results from network analysis, local nonbonded interaction energies, and free energies of binding all show that the Pre-transfer and Post (H-AMP) states form stronger GluRS·tRNA interactions than all other post-transfer states regardless of the presence or absence of AMP. The pKa calculations suggest that one of the α-ammonium hydrogens on the charging glutamate can transfer to the Glu41 sidechain carboxylate while in the active site. Glu41 is predicted to be a nearly universal handle among Class I aaRSs that acts as a general base to facilitate tRNA release from the active site upon transfer of the amino acid. If Glu41 becomes solvated, either through the removal of AMP or the dissociation of the CCA hairpin, it would return to its charged state. Similarly, the pKa measured at the beginning and end of the Post (no AMP/GluNH_2) simulation shows that once the CCA end leaves the active site, breaking contacts with Glu41, the α-amino group on the charging glutamate can become reprotonated. Principal component analyses of the tRNA in both uncharged and charged states indicate that charged tRNA motion is highly coupled in all regions as opposed to less correlated when the tRNA is uncharged, which may assist in tRNA undocking immediately following aminoacylation. From binding free energies it appears that the pathways for AMP and tRNA dissociation are thermodynamically competitive, but once AMP has left the active site and the α-ammonium group is deprotonated, dissociation of the CCA hairpin can occur on the nanosecond timescale. Binding of EF-Tu to the GluRS·$tRNA^{Glu}$ can stimulate tRNA release. Further studies will be needed to determine the molecular details of the migration of the tRNA to the EF-Tu and whether its complete dissociation from the aaRS occurs prior to or during delivery of the tRNA to the ribosome A-site.

3.6 Acknowledgements

The authors thank ZLS group members, particularly Li Li and Elijah Roberts, for many helpful discussions. They also wish to thank Nathan Baker for APBS assistance, Jan Jensen for help with PROPKA 2.0, Susan Martinis for experimental interpretations, and John Stone for VMD graphics suggestions. Funding for ABP, JE, and AS was provided by NSF (MCB04-46227), NSF (MCB08-44670), NSF (PHY08-22613), and NIH Chemical Biology Training Grant (5T32GM070421). Supercomputer and local computing time was provided by NCSA LRAC (MCA03T027) and and NSF CRIF (0541659).

CHAPTER 4

MODELING AND DYNAMICS OF THE CYSRS·TRNA^{CYS} COMPLEX

4.1 Summary

CysRS is the enzyme that loads tRNA^{Cys} with cysteine in most bacterial/eukaryal organisms and nonmethanogenic archaea. Following landmark studies of the indirect pathway for tRNA^{Cys} charging in methanogenic archaea [139, 384], the next step was to explore molecular details of charging and dissociation of tRNA^{Cys} in the direct pathway. In this study, an active structure of CysRS·tRNA^{Cys} from *E. coli* was modeled from the inactive crystal structure [8]. Through alignment of homologous structures, the docking interface of the CysRS crystal structure was determined to be significantly shifted and multiple modeling techniques were employed to build a structure that was correctly docked. Molecular dynamics simulations of the model and subsequent energetic and bioinformatics analyses demonstrated that the model was correct on a coarse-grain level. However, the estimates of the binding free energy showed that the structure was unstable due to absent interfacial interactions. The model is therefore sufficiently accurate to support other computational studies of aaRS·tRNA binding, but cannot be used for dynamics studies [1].

[1]The model was published in the supporting information in Alexis Black Pyrkosz, John Eargle, Anurag Sethi, and Zaida Luthey-Schulten. "Exit strategies for charged tRNA from GluRS," *J Mol. Biol.*, 397, 1350-1371 © Elsevier 2010. All modeling, simulations, and analysis were performed by ABP. Analysis of Class I a/b structurally conserved interactions and parameterization of ms^{2}i^{6}A were done by Anurag Sethi.

4.2 Introduction

As stated in Section 3.2, the aminoacyl-tRNA synthetases (aaRSs) charge their cognate tRNAs with the correct amino acids, assisting in maintenance of the genetic code. Once charged, tRNA is shuttled to the ribosome by elongation factor Tu (EF-Tu). As shown in Figure 3.1, charging of tRNA by the aaRSs occurs in several several stages, from recognition and binding of the correct amino acid and tRNA from the cellular pool, formation of the adenylate, transfer of the amino acid to tRNA, and release of charged tRNA to the waiting EF-Tu. While this figure shows the pathway specific to those Class I aaRSs requiring presence of tRNA to form the adenylate, in most aaRSs this reaction occurs prior to the docking of tRNA. As also stated in Chapter 3, texit strategies available for the undocking of charged tRNA from Class I aaRSs are many and experimentally unverified. Stimulation of the charged tRNA undocking from the aaRS by the presence of EF-Tu has been verified both experimentally and computationally [76, 78, 132]. The hypothesis is that EF-Tu docks to the aaRS·tRNA complex, causing the 3′ end of the tRNA to exit the aaRS active site and dock into EF-Tu while the anticodon remains bound to the aaRS [315]. To form a model of the quaternary complex would require knowledge of precise molecular details of tRNA docked to both aaRS and EF-Tu from structural studies. Currently available crystal structures suggest that this study could be performed in only one Class I system: cysteine [100, 8].

Background and Phylogenetic History of the CysRS·tRNA^{Cys} As stated previously, aaRSs are divided into two classes based on the catalytic domain and Class I aaRSs have a Rossman fold. The active site is located in C-terminal loops of the β-strands and is flanked by conserved HIGH and KMSK sequences that bind ATP during adenylate formation. The connective polypeptide (CP1) insertion is located between the two halves of the Rossman fold and closes over the 3′ end of the tRNA during aminoacylation. CysRS is nominally part of the Class Ia subclass with a small CP1 insertion and a set of parallel α-helices (four-helix bundle or 4HB) connecting the catalytic domain (CD) to the

C-terminal domain [8]. This C-terminal domain (anticodon binding domain or ACB) is a mixed α/β fold unique to CysRS and directly binds the $tRNA^{Cys}$ anticodon [8, 131]. Characteristic of CysRS is the presence of a Zn^{2+} ion in the active site. This ion has been shown to act as a pre-editing mechanism by screening the charging amino acid prior to

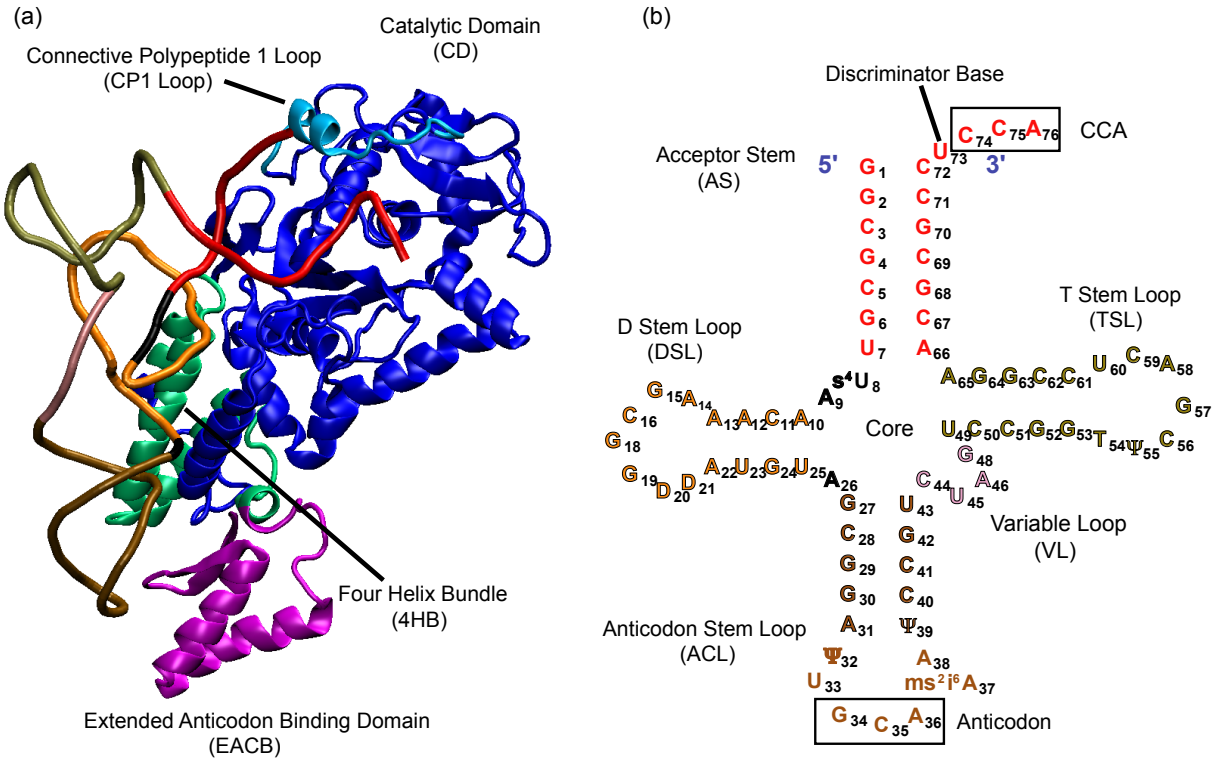


Figure 4.1: Structure and sequence of the modeled CysRS·$tRNA^{Cys}$. Panel (a) shows CysRS with functional domains labeled while (b) contains the $tRNA^{Cys}$ cloverleaf schematic where nucleotides are labeled with their canonical numbering and colored according to their structural region in (a). Core nucleotides are outlined in black.

$tRNA^{Cys}$ docking [385]. The thiol on the sidechain of the charging cysteine coordinates to the Zn^{2+} ion, which is similar to the GluRS system shown in Chapter 3 where the charging glutamate sidechain is bound by conserved arginine residues in the active site. CysRS has been crystallized in two states, in complex with charging cysteine or $tRNA^{Cys}$ [344, 8]. In the first structure, the CysRS from *Escherichia coli* contains the charging cysteine coordinated to the Zn^{2+} and is missing residues in the CP1 insertion and ACB. The second structure was used for the current study and contains *E. coli* CysRS with its cognate

transcribed $tRNA^{Cys}$ in the active site [PDB code: 1U0B] [8]. The tRNA is unproductively bound because the last nucleotide in the system, A76, which is the site of amino acid charging, is docked into the adenylate binding site. This structure is used as a starting point for this study wherein the tRNA is modeled to dock in a productive manner and the active site contains the adenylate (see Figure 4.1). CysRS has also been the subject of extensive biochemical studies performed to characterize interface binding and identity elements [155, 189, 190, 191, 192, 194, 193, 195, 198, 196, 199, 5, 197].

CysRS has a noncanoncial evolutionary history [64, 139, 91, 384, 386] wherein the CysRS enzyme was originally of bacterial origin and horizontally transferred in multiple independent events to eukaryal and archaeal organisms. Methanogenic archaea were recently shown to use an indirect pathway for charging $tRNA^{Cys}$ with cysteine by using a Class II type aaRS (SepRS) to load *O*-phosphoserine, the precursor to cysteine along the biosynthetic pathway, on $tRNA^{Cys}$. The precursor would be converted to cysteine by the enzyme SepCysS. Through horizontal gene transfer, several archaea acquired the bacterial type CysRS and the modern archaeal organisms contain a mixture of enzymes with some organisms using SepRS/SepCysS exclusively, some relying completely on CysRS, and others using them jointly to various degrees. Because bacterial CysRS contains regions that are common to all three domains of life, the current study employs bioinformatic analyses over all three domains of life as represented by the *E. coli* CysRS.

Modeling the active CysRS·$tRNA^{Cys}$ structure The main goal of this study is to use the current unproductive CysRS·$tRNA^{Cys}$ structure to model the active complex in the pre-transfer state. According to the opinion of the field, this modeling is expected to consist largely of modifying the angles of the backbone phosphates of the $3'$ end of the tRNA to position A76 for nucleophilic attack on the adenylate. Instead, we use comparative structural analysis of homologous Class I aaRSs to show that crystallization artifacts have produced a structure with a protein·RNA interface that significantly differs from that of the rest of the class. Further, we combine interactive molecular dynamics and homology modeling techniques to produce a structure of the CysRS·$tRNA^{Cys}$·Cys-AMP

that is consistent with homologous structures. We validate the model using molecular dynamics, evolutionary analysis and local energetics to identify key interface contacts, and free energy of binding estimates to evaluate complex stability over 20 ns of simulation.

4.3 Methods

4.3.1 Bioinformatics

Protein All CysRS bacterial sequences in the Swiss-Prot database [350] were downloaded and aligned with ClustalW [149] in the MultiSeq plugin [150] to VMD [151]. The alignment was improved manually. SeqQR was applied at a 65% cutoff [387] to obtain the final evolutionary profile over all three domains of life with the *E. coli* sequence used to seed the profile.

TRNA The full evolutionary profile for $tRNA^{Cys}$ was obtained by downloading all $tRNA^{Cys}$ sequences from the tRNA Compilation 2000 [341] and the Integrated Microbial Genome database at the Joint Genome Institute [145]. The sequences were aligned using ClustalW [149] and improved manually. The SeqQR algorithm was used with a 85% cutoff with norm order = 3 and gap penalty = 0.5 to produce the nonredundant set over all three domains of life.

4.3.2 Molecular modeling

An examination of the aaRS:tRNA binding interface in other Class I crystal structures was conducted by structurally overlapping the complexes from the Glu [1N78] [316], Gln [1QTQ] [342], Arg [1F7U] [343], and Leu [1WZ2] [388] systems according to the catalytic domain (SCOP definitions) [389] with the MultiSeq plugin [150] in VMD [151]. The alignment showed excellent agreement across these systems, but poor overlap with the inactive Cys structure (see Figure 4.2 in Results). To correct this, the system was

completed by adding missing sidechains and the missing 71-76 residue stretch with Modeller v6.2 [390] and with hydrogens, histidine protonation states, ions, and solvent as will be explained below in preparation for the next step. Modified bases were also added, their positions and parameters derived from the EF-Tu:tRNA^{Cys} structure 1B23 [100] and subsequent computational study [63]. The interactive molecular dynamics (IMD) plugin [391] in VMD was used to place forces on the tRNA backbone and shift the conformation to more precisely match the other Class I structures. The structurally conserved contacts in the multi-aaRS overlap (D71-C75, K73-A76, K185-A76, and W205-A76) were used to evaluate the placement of the CCA hairpin, and sidechains were rotated according to the 1LI7 structure to create these contacts. Upon completion, the CCA hairpin backbone angles was changed to match tRNA^{Gln} in 1QTQ. The charged cysteine and Zn^{2+} ion were placed according to 1LI7 [344], which contains the free CysRS with the charged cysteine. The AMP was modeled by analogy to the other Class I structures containing AMP or adenylate analogs. Because the pKa of cysteine in the presence of Zn^{2+} ions is 8.5, the thiols on the charged cysteine, C28, and C209 were deprotonated [392]. The parameters for the deprotonated cysteine were developed using Hartree Fock 6-31G* in Spartan05 and the CHARMM27 force field [355]. The Cys-AMP was formed analogously to the Glu-AMP. To ensure that the modeled portions of the system were stable, the system was minimized and equilibrated for 1 ns as detailed below.

Histidine protonation states were determined by using the Whatif server [351] and visual inspection. The system was neutralized by placing Mg^{2+} and K^{+} ions with the program Ionize [2], which calculates the Coulombic interaction energies for a system based on a uniform grid and places ions at the most energetically favorable points. The Mg^{2+}placement protocol developed previously was used to place 3 Mg^{2+} at 2 Å, 14 Mg^{2+} at 6.5 Å, and the remaining charges as K^{+} at 6.5 Å away from the macromolecules. The Mg^{2+} ion primary solvation shells were completed with TIP3 water molecules [352] . To ensure full solvation of the active site and binding interface, Solvate 1.0[3] was used with two

[2]http://www.ks.uiuc.edu/Development/MDTools/ionize

[3]Grubmueller, H. 1996 SOLVATE 1.0. http://www.mpibpc.gwdg.de/abteilungen/071/solvate/docu.html

gaussians to add two layers of TIP3 [352] water molecules to the system. Solvate 1.2 in VMD was used to complete the water box, reaching an approximate system size of 120 x 80 x 120 Å and 100,000 atoms.

4.3.3 Molecular dynamics

The partial atomic charges for AMP were derived based on ATP (available in CHARMM27). Parameters for the bond between A76 and cysteine were taken by analogy from the *O*–phosphoserine system [384].

The parameters for the modified bases were derived by analogy the respective transcribed bases and all charges and equilibrium bond lengths and angles were verified using *ab initio* quantum chemistry calculations with the 6-31G* basis set in Spartan04 (Wavefunction, Inc., Irvine, CA).

Dihydrouridine and 4-thiouridine The parameters for dihydrouridine were available as a patch in the Unified Atom force field [393], which is distributed as a stream file with the CHARMM27 force field. Hydrogens were added to this patch by analogy with uridine and cyclopentane. The parameters for 4-thiouridine were derived by analogy with uridine.

Pseudouridine The parameters for the standard Ψ tautomer were developed by analogy with cytidine and showed agreement with parameters provided by Lennart Nilsson. These were used for Ψ39 and Ψ55. A different tautomer was used for Ψ32 based on previous simulations of $tRNA^{Cys}$ docked with EF-Tu [63], in which after roughly 1 ns of simulation, the base destacked from the anticodon loop and formed a hydrogen bond with the A31 phosphate. The anticodon arm would then unravel. The Ψ32 tautomer had the N3 hydrogen moved to O4, forming a hydrogen bond with its own phosphate, and retaining the crystal structure conformation. The parameters were derived by analogy to uridine and cytidine as well as a guanine tautomer in a stream file.

2-methylthio-N^6-isopentenyladenosine The parameters for ms^{2}i^{6}A were derived by analogy with adenine, ethyl methyl sulfide, ethenyl methyl thioether, 2-butene, dimethyl amine, and isopentenyl amine. All quantum chemistry calculations for the parameterization of Mia were performed with GAMESS [394] using the MP2/Hartree Fock methods with the 6-31G* basis set. The force constants for all bond parameters (bond stretching, angle bending, and torsion) were obtained by scaling the corresponding terms from the quantum chemistry calculations with similar CHARMM27 parameters from the molecules listed above. The equilibrium values for these bonding terms were obtained from the optimization step of the quantum chemistry calculations.

All parameters were tested using a short minimization and equilibration with NAMD2. The parameters were also validated by comparing the conformations of the corresponding nucleotides in the tRNA simulation with those in the crystal structure.

4.3.4 System setup

Simulation was performed with the molecular dynamics software NAMD2 [354], using the NPT ensemble with periodic boundary conditions, long range electrostatics calculated with Particle Mesh Ewald summation [357], and pressure set to 1.01325 bar. The timesteps were 1 fs for bonded, 2 fs for VdW, and 4 fs for electrostatic force calculations. A cutoff of 12 Å and switching distance of 10 Å was used for the VdW force calculations. Snapshots were saved every 500 fs.

Minimization was carried out in steps to ensure that the small molecule substrate in the active site was stable. The first 50,000 steps were performed in 5,000 step increments to relax different portions of the modeled structure: 1) all hydrogen atoms, 2) all hydrogen atoms, water molecules, and Mg^{2+}/K^+ ions, 3) the discriminator base (U73), 4) all hydrogen atoms, water molecules, and the CCA hairpin, 5) all hydrogen atoms, water molecules, and all active residues with 4 Å of the adenylate, 6) all hydrogen atoms, water molecules, and the adenylate, 7) all hydrogen atoms, water molecules, and active residues bound to the Zn^{2+} ion, 8) all hydrogen atoms, water molecules, ions, and amino acid

sidechains, 9) all hydrogen atoms, water molecules, ions, and the tRNA bases (except the first base pair of the acceptor stem to prevent unzipping due to modeling artifacts), 10) all nonbackbone atoms, and 11) all nontRNA backbone atoms. The last step was run for 20,000 steps of unconstrained minimization. The model was evaluated according to the structurally conserved contacts listed previously (see Molecular Modeling section) to ensure that the maximum number of contacts were preserved. The position of the adenylate was compared to the multi-aaRS overlap and the coordination of the charged cysteine to the Zn^{2+} ion was compared to the 1LI7 crystal structure for stability.

Equilibration was conducted using a temperature jump protocol developed by Auffinger and Westhof [358]. By systematically raising the temperature and gradually freeing of harmonic constraints, the Mg^{2+} and K^{+} ions were able to enter the deep grooves of the highly negatively charged tRNA. The steps were: 1) a temperature of 100 K with only hydrogens unconstrained for 25,000 fs, 2) a temperature of 200 K with hydrogens and waters freed for 25,000 fs, 3) a temperature of 250 K with all hydrogens, waters, ions, and ligands unconstrained for 25,000 fs, 4) the temperature held at 250 K and all nonbackbone atoms freed for 25,000 fs, and 5) a final temperature of 298.15 K and all atoms were unconstrained for 19 ns. The harmonic constraints in all steps were set to $1\,\mathrm{kcal\,mol^{-1}\,\AA^2}$.

4.3.5 RMSD calculations

To ensure that each system was equilibrated before gathering statistical data, the RMSD for the C^{α} and P atoms in the aaRS and tRNA was calculated over time by structural overlap of frames at 10 ps intervals with the initial structure. Because the system showed a leveling of RMSD after 4 ns, the last 16 ns of the trajectory was used for further analysis.

4.3.6 Local energetics analysis

To calculate the local energetics, the NAMDEnergy plugin in VMD was used with CHARMM27 force field parameters. The non-bonded energy terms (VdW and

electrostatics) were calculated between each nucleotide with the protein or each residue with the tRNA in frames taken from the last 16 ns of the 20 ns equilibration at a frequency of 10 ps. A switching distance of 18 Å and a cutoff of 21 Å was used to ensure that protein contacts were calculated relative to both members of a base pair (base pairs average 18 Å across). These values were averaged over all frames. A mask was calculated based on the multiple sequence alignment of each molecule (see Bioinformatics in Methods). For CysRS and $tRNA^{Cys}$, the percent sequence identity over sequences from all three domains of life was determined based on the *E. coli* sequences. Modified bases were not taken into account due to the lack of data for all sequences. An energy cutoff of ±100 and ± 50 kcal/mol was used on aaRS and tRNA interaction maps respectively.

4.3.7 Free energies of binding

The Molecular Mechanics–Poisson-Boltzmann Surface Area (MM-PBSA) method is used here as in Chapter 3 to calculate binding free energies from a simple thermodynamic cycle [365, 366, 367, 368, 369, 370, 63, 371]. The difference in energy between the complex and two components or unbound docking partners is calculated according to $\langle \Delta G_{binding} \rangle = \langle \Delta E_{VdW} + \Delta E_{coul} + \Delta G_{nonpolar} + \Delta G_{polar} \rangle - T\langle \Delta S \rangle$. Each averaged value is the difference for tRNA dissociating from the CysRS·Cys-AMP complex ($\langle \Delta G^{tRNA}_{binding} \rangle = \langle G_{complex} - G_{CysRS \cdot Cys-AMP} - G_{tRNA} \rangle$). The three bound Mg^{2+} ions placed during system setup and water molecules in their first solvation shell were included with the tRNA, which bind the tRNA core and are stable elements of the nucleic acid structure. Calculations for individual components were performed using the same trajectory as the complex [367, 63], neglecting contributions from conformational changes in individual components.

The bonded interaction differences (bonds, angles, dihedrals, and impropers) cancel in each calculation and therefore are not shown. The E_{VdW} term is the sum of pairwise VdW interaction energies and was calculated using the NamdEnergy plugin in VMD. The Coulombic energy (ΔE_{coul}) is the sum of all pairwise interactions in the system scaled by ϵ_{in}, was calculated with Coulomb, a program distributed with APBS [372].

The nonpolar solvation free energy ($\Delta G_{nonpolar}$) is the energetic cost of creating a cavity in the solvent for a given system. $G_{nonpolar} = \gamma SASA + b$ was calculated for the complex, CysRS·Cys-AMP, and $tRNA^{Cys}$ with SASA set to the solvent accessible surface area using a solvent radius of 1.4 Å, $\gamma = 0.00542\,\text{kcal mol}^{-1}\ \text{Å}^{-2}$, and $b = 0.92\,\text{kcal/mol}$. This term was calculated in VMD. The polar solvation (ΔG_{polar}), the energy required to move charged solute from vacuum dielectric ($\epsilon_{in} = 1.00$) to aqueous dielectric ($\epsilon_{out} = 78.54$), was computed for the complex and individual components. The 1.00 interior dielectric was selected based on calculations of $tRNA^{Glu}$ binding to GluRS in Chapter 3. This term was numerically calculated with the Poisson-Boltzmann solver APBS [372]. A temperature of 298.15 K and CHARMM27 radii/charges were used on a grid of 193 x 129 x 193 with automated focussing. Grid size was based on the crystal structure with 15% padding in each dimension. The implicit salt concentrations of 117 mM KCl and 38 mM $MgCl_2$ were used to model the explicit concentrations in the system states. Each frame was aligned to the modeled structure based on the C^{α} and P atoms to limit error from orientation changes. The grid was centered on the complex.

The entropy terms (ΔS) were estimated from the covariance matrix of atomic position fluctuation using Schlitter's formula [373, 374, 375], which relates the entropy of a solute molecule to the sum of decoupled simple harmonic oscillators obtained from the principal component modes in the molecular dynamics simulation. The coordinates of the backbone atoms (C^{α} and phosphorous) intervals of 10,000 frames from the 5 ns of the equilibration trajectory were used to generate the mass-weighted covariance matrix; the program Carma [361] was used to calculate the eigenvalues. These were then used to calculate the determinant and substituted into Schlitter's formula to calculate the entropy of the complex and individual components [63].

4.4 Results and Discussion

4.4.1 Molecular modeling of CysRS·tRNA^{Cys} Complex

Model of active complex The first goal was to model the unproductive crystal structure into a functional complex. Through structural comparison of all Class I a/b crystal structures containing aaRS·tRNA, the binding site of A76, adenylate, and charging amino acid were determined to be nearly identical. However, changing the backbone angles of the tRNA^{Cys} CCA hairpin nucleotides to place A76 in a functional location was unsuccessful; the length of the hairpin was consistently one nucleotide too short to reach the A76 binding site. This indicated that greater conformational changes were required and therefore the structure is more inaccurate than previous anticipated.

To characterize these conformational changes, the CysRS crystal structure was added to the Class I a/b crystal structure alignment. The alignment was based on the catalytic domain, which is similar in all Class I aaRSs. As shown in Figure 4.2, GluRS, GlnRS, ArgRS, and LeuRS structures in complex with their respective tRNAs show good agreement in both protein structure alignment and docking interface conformations. CysRS aligns well with the other aaRS structures, but the docked conformation of tRNA^{Cys} differs significantly from the others. The core (D/T arms) is shifted further away from the interface. The acceptor stem reaches toward the active site from a more obtuse angle (the angle formed at the elbow region) and is unable to fully span the distance to the A76 binding site. While it is conceivable that CysRS docks tRNA^{Cys} in a different conformation than used in other Class I aaRSs, the fact that the CCA hairpin cannot be moved in a trivial manner to reach the charging cysteine (whose binding position was determined previously in a crystal structure without tRNA^{Cys} and is strongly coordinated to the Zn^{2+} ion) indicates that the entire interface has been captured in an unproductive manner and requires significant modeling to correct.

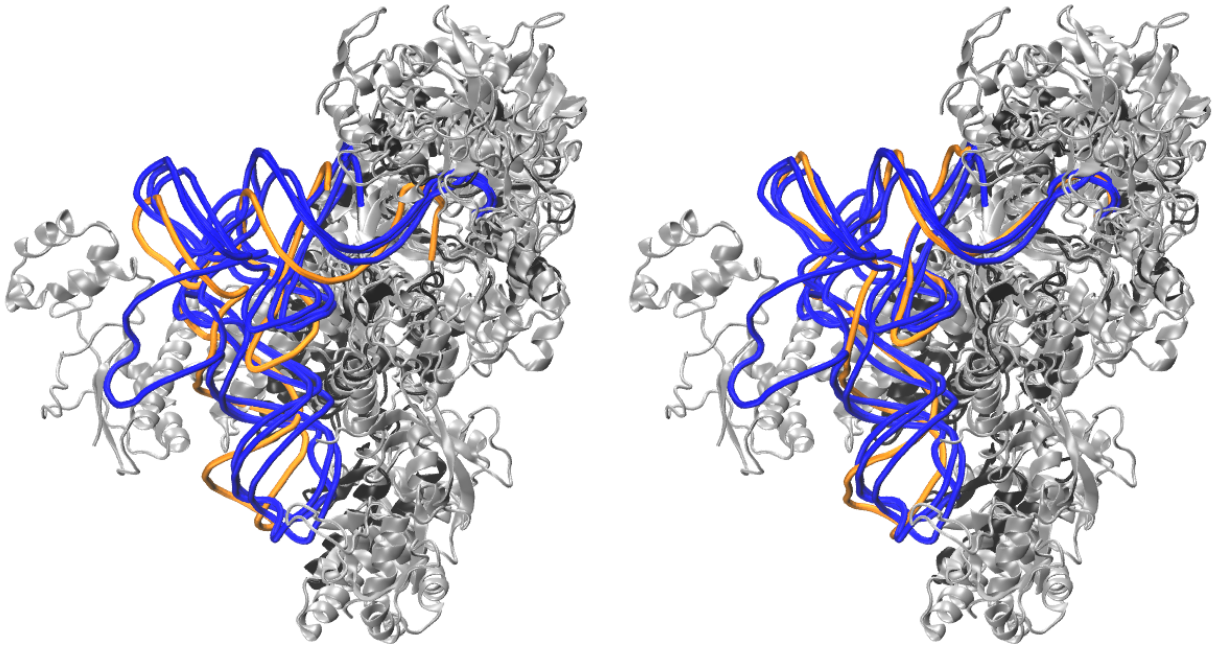


Figure 4.2: CysRS·tRNA^{Cys} before and after modeling. In panel (a), the crystal structures of the other Class I aaRSs in productive conformations are shown in blue (tRNA) and silver (protein) while the CysRS·tRNA^{Cys} crystal structure is shown in orange (tRNA) and black (CysRS). The structures were aligned by the catalytic domain. In panel (b), the same alignment is shown, but with the crystal structure of the CysRS·tRNA^{Cys} complex replaced by the model.

The modeling was performed using a combination of interactive molecular dynamics (IMD), automated homology modeling, and local manual homology modeling. IMD is a computational technique that permits manipulation of molecules in molecular dynamics simulations with real-time force feedback and graphical display [391] and was used here to move the tRNA^{Cys} backbone into a conformation similar to the other Class I tRNAs. Automated homology modeling was used to return the active site residue conformations following IMD back to the crystal structure. Finally, local manual homology modeling was used to adjust the backbone angles of the CCA hairpin and the first base pair of the acceptor stem to match those of the tRNA^{Gln}, one of the most closely homologous systems.

The final CCA hairpin position placed A76 in its determined binding site, close to the charging cysteine. Figure 4.2b shows the final structure of the CysRS·tRNA^{Cys} model.

To validate the model, a 5 ns molecular dynamics simulation was evaluated for key interactions in the active site. The structural alignment of the Class I a/b aaRSs showed several residue·nucleotide contacts that were conserved and similar interaction pairs were found in the simulation of the modeled structure.

Conserved Active Site Interactions	
Interaction Pair in CysRS	Homology
Thr31 hydrogen bonds A76 2′ OH	GluRS: Ser9-A76, GlnRS: Asp34-A76, ArgRS: Asn153-A76
Thr68 hydrogen bonds to A76 3′ OH	GluRS: Asp41-A76, GlnRS: Asp66-A76
Asp71 near C75 2′ OH	GluRS: Asp44-C75, GlnRS: Asn69-C75
Lys73 near CCA hairpin	GluRS: Arg47-C75, ArgRS: Gln195-A76
Lys185 salt bridge A76 phosphate	GlnRS: Arg192-A76, ArgRS: Lys340-A76
Trp205 base stacks A76	GluRS: Tyr187-A76, GlnRS: Tyr21-A76, ArgRS: Tyr347-A76

Table 4.1: Structually conserved active site interactions between residues and nucleotides

4.4.2 Protein·tRNA interface contains highly conserved interactions

tRNA^{Cys} nucleotides forming interfacial interactions are highly conserved To identify those nucleotides in tRNA^{Cys} that make important interface contacts, a bioinformatics study pinpointing highly conserved nucleotides across all three domains of life was conducted. There are two sets: the first is composed of the universally conserved U8, A14, G15, G18, G19, U33, G52, G53, G54, G55, C56, A58, U60, and C61, while the second contains those that are specific to tRNA^{Cys}: G1, C11, A22, G24, G30, G34, C35, A36, A38, C40, C72 and U73. These are 85% conserved or greater and are predominantly in three regions of the tRNA: the core, the anticodon loop, and the acceptor stem. The

C11·G24 and G30·C40 base pairs are in the D arm and anticodon stem respectively and form the strongest base pairs in their stems. G24 and C40 are also at the interface, forming backbone contacts with the sidechains of CysRS residues Gln311 (20%), Arg318 (48%), Arg364 (28%) and Arg318, Arg364, Asn367 (59%) respectively. These CysRS residues are all located in the 4HB and collectively form a pocket of positive charge near the junction between the 4HB and EACB (see Figure 4.3 for a structural reference). Their moderate conservation is deceptive because the 4HB is highly variable, even among closely related bacteria. The lack of crystal structures from organisms other than *E. coli* prevents further investigation, but the length of the helices and the spacing between them most likely varies, preventing accurate alignment based solely on sequences.

The anticodon nucleotides G34, C35, and A36 are completely conserved in $tRNA^{Cys}$. Despite cysteine having two codons, UGU and UGC, which are both used to varying degrees in a variety of organisms[395], only tRNA sequences containing the GCA anticodon have been sequenced to date for cytosolic $tRNA^{Cys}$[341, 340, 135]. Of the three nucleotides, G34 is the most strongly bound to CysRS and the base makes two hydrogen bonds each to the sidechains of Arg423 (69%) and Asp436 (86%) as well as stacking with the Trp432 sidechain (17%). Trp432 is not conserved but only aromatic amino acids capable of forming the base stacking interaction are found in this part of the CysRS alignment from among the other organisms. This highly specific hydrogen bond network indicates that only guanine is recognized in this position, not the adenine that would bind the alternative codon. Guanine is known to form a wobble base pair with uracil at the ribosome[114]. C35 base makes hydrogen bonds with the sidechains of Arg439 (82%) and Asp451 (72%). A36 base makes a comparatively minimal interaction with the backbone of Asp451 as observed previously[8], but has been shown by biochemical assays to be essential for aminoacylation[155]. All of these residues are conserved and located within the ACB.

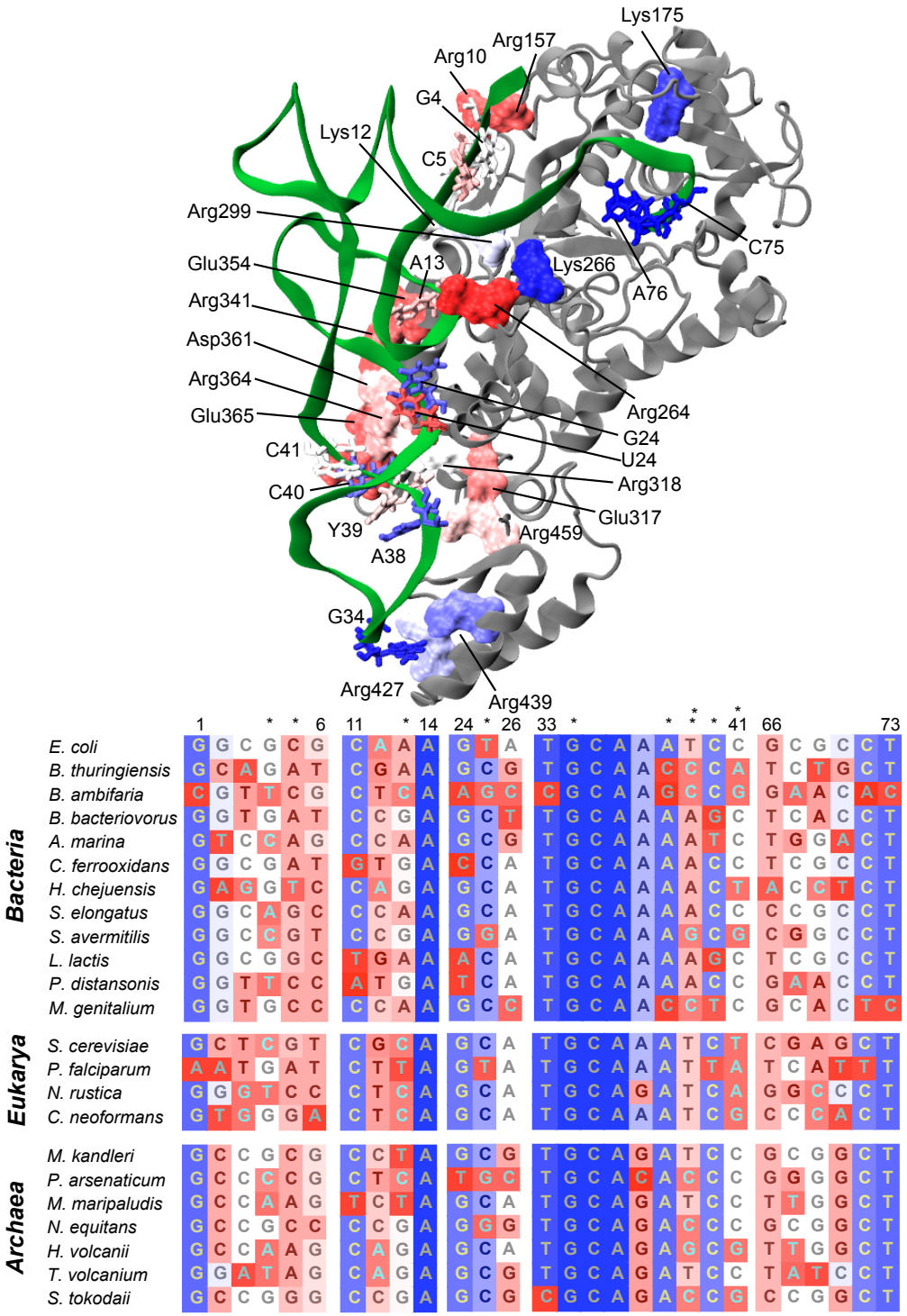


Figure 4.3: CysRS·tRNA^{Cys} with conserved interfacial contacts. tRNA^{Cys} nucleotides and CysRS residues making large energetic contributions to binding are shown in licorice and surfaces, respectively. Color is based on percent sequence identity from the evolutionary profile over all three domains of life (a subset is shown for tRNA^{Cys}) with blue, white, and red showing high, moderate, and poor conservation, respectively.

U73, the discriminator base, is highly conserved, but does not make strong interactions with any CysRS residues. U73 is found only in $tRNA^{Gln}$ (eukaryotic), $tRNA^{Gly}$ (bacterial), $tRNA^{Thr}$ (archaeal and bacterial), and all three domains of life for $tRNA^{Cys}$ [63], with crystal structures of ThrRS (archaeal [396]/bacterial [397]) and CysRS (bacterial [344]) containing Zn^{2+} in the active site (structures for the other aaRSs in the indicated domain of life are unavailable). Due to the bent hairpin conformation necessary for A76 to approach the charging cysteine as it coordinates to the Zn^{2+} in CysRS, U73 may be necessary to stabilize the required destacked and contorted 3' end of the tRNA in aaRSs containing this metal ion. This theory can be investigated further when more crystal structures become available. The rest of the CCA hairpin is in contact with few amino acids in the active site of CysRS. In the modeled structure, the conserved Gly29, Thr33, and Thr68 form hydrogen bonds with the α–ammonium group of the charging cysteine.

The modified bases in $tRNA^{Cys}$ in *E. coli* are located in positions that are either conserved or predominantly one of two bases. U8, 4-thiouridine, is nearly universally conserved. The dihydrouridines in the D loop are generally in positions 20 and 20a (canonical numbering) after the universally conserved G18 and G19 in bacteria and eukarya, but are also known to precede these nucleotides (positions 17 and 17A). Position 32 is either a cytosine or a pseudouridine, indicating that the amine as located on cytosine is necessary to maintain the anticodon loop structure. Position 37 often contains a modified adenine, as in *E. coli*, or a methylated guanine, which is more common in eukaryotic sequences. The modified uracil at position 54 and pseudouridine at position 55 are universally present, whether 54 is a thymine or 1-methyl-pseudouridine (*Bacteria*/*Eukarya* and *Archaea* respectively). One of the two modified bases present at the interface, ms^2i^6A37, has the bulky allyl chain pointing away from CysRS. The highly conserved adjacent unmodified nucleotide A38 may be necessary to complement the presence of this hypermodified base. The other modified base at the interface, $\Psi 39$, forms a base pair with A31 with the same hydrogen bond pattern as a uracil. This position is not conserved.

CysRS interface There are a number of interactions between the acceptor stem with residues in the CP1 insertion, which are not highly conserved. In this region, S156 is partially conserved in the bacterial and archaeal sequences while it is not conserved in the eukaryotic sequences, and R168 is partially conserved in all domains of life and these residues form hydrogen bonds with the ribose of C72 and the phosphate of G70 respectively. While C72 is highly conserved across tRNAs of all specificities and all domains of life[63], G70 is not. However, as the interaction of S156 with G70 is backbone mediated, this may not put a constraint on the base of this nucleotide and is a shape specific interaction. Similarly, C5 forms a salt bridge with the partially conserved R10. The CP1 loop is variable both within and across all domains of life, but usually contains many hydrophilic amino acids. Also, the eukaryotic sequences showed an insertion in the CD of nearly 100 residues near residue 80 which occurs at the loop following β–strand 2 in the Rossman fold, indicating that there could be an additional subdomain, possibly an editing domain, in the eukaryotic system and as such, the interactions of this region with the tRNA may not be conserved across all domains of life.

4.4.3 Modeled Cys-AMP is stable in the CysRS active site

A molecular dynamics simulation was performed on the CysRS·tRNA^{Cys} in the Pre-transfer state for 20 ns at 298 K and 1 bar (see Methods). The following sections use this simulation and comparisons to those run on GluRS·tRNA^{Glu} (see Chapter 3) to evaluate the modeled CysRS·tRNA^{Cys} complex for its structural and interfacial accuracy. To accurately sample only equilibrated parts of the trajectories, the RMSD relative to the starting structure was used to monitor the equilibration with the initial burst of rising RMSD leveling off around 4 ns. Therefore, subsequent data processing was performed on the final 16 ns of the trajectory.

The primary goal of modeling the CysRS·tRNA^{Cys} structure was to produce a structure with a productive active site. In the original crystal structure, A76 was docked into the AMP binding site and the adenylate was absent. In the modeled structure, the

aminoacyl-adenylate was placed by analogy to the other Class I structures containing adenylates or amino acids and A76 was oriented so the 2' hydroxyl group could nucleophilically attack the carbonyl carbon on the adenylate. During equilibration, the active site residues in the adenylate binding site formed numerous hydrogen bonds and salt bridges with the adenylate (see Figure 4.4). The conserved triad residues binding the α–ammonium group of the adenylate described in Chapter 3 are shown here to be Gly29 (backbone), Thr31 (sidechain), and Thr68 (sidechain). Thr68 is structurally analogous to Glu41 in GluRS, but does not form a salt bridge, presumably due to the close proximity of the Zn^{2+} ion. The negatively charged sulfur atom on the charging cysteine is strongly bound to the Zn^{2+}, as in the crystal structure containing CysRS and cysteine [344]. The Zn^{2+} ion retains the trigonal bipyramidal geometry observed in both crystal structures, though the exact distances shift slightly from the 1U0B structure with no adenylate to the modeled structure with adenylate. The distances between the other active site residues and the adenylate also indicate that the residues have shifted position slightly to accommodate the presence of the adenylate. However, the distances are not as small as in the GluRS active site (see Figure 3.6). In GluRS, Leu236 tightly binds the adenine base of the adenylate, while in CysRS, Glu258 only loosely associates with it. Also in GluRS, Ser9 interacts with the phosphate with an average distance of 2.82 Å, while Thr31 in CysRS binds the phosphate at a distance of 3.91 Å. Also, in GluRS His15 in the HIGH motif binds the AMP phosphate only after the tRNA has been charged, while in CysRS His40 in the HIGH motif binds the adenylate ribose throughout the Pre-transfer simulation. Slight differences in adenylate binding between aaRSs are expected; a general pattern of binding is seen between the adenylate and active site residues in the CysRS and GluRS structures. The active site of the modeled structure is sufficiently similar to the GluRS structure to make general statements about the binding residues. However, the weaker interactions indicated by the longer distances suggest that the model may lack accuracy, which may affect the binding energetics.

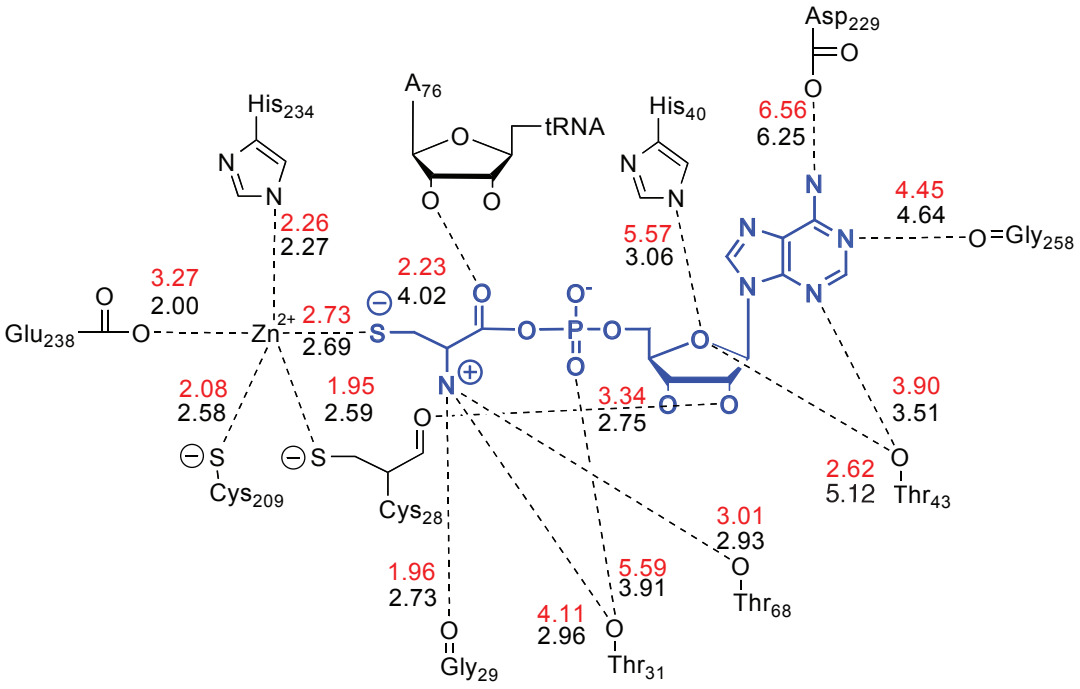


Figure 4.4: Interactions between the aminoacyl-adenylate and active site residues. The adenylate is shown in blue with dotted lines indicating interactions to CysRS residues, A76 on the tRNA, or the Zn^{2+} ion. Distances between the heavy atoms are labeled in red for the modeled structure and in black when averaged over the last 16 ns of the 20 ns equilibration. All hydrogens have been removed for clarity.

4.4.4 Stabilizing interactions are missing in the model

AaRSs specifically recognize and bind their cognate tRNAs through contacts spanning the acceptor stem to the anticodon, often through identity elements. To ascertain if the CysRS·$tRNA^{Cys}$ model was accurate, the contacts making substantial energetic contributions between identity elements and other conserved nucleotides in $tRNA^{Cys}$ to conserved CysRS residues were identified by averaging the electrostatic and van der Waals interaction energies over the last 16 ns of the Pre-transfer simulation. As a comparison, similar interactions were identified from simulations of the GluRS·$tRNA^{Glu}$ in the presence of Glu-AMP, and these are shown in Figure 4.5. The energies were calculated with a cutoff of 21 Å to include both sides of the bound tRNA helices while limiting the analysis to the aaRS binding interface. The conservation of the nucleotides/residues is shown by masking

the energies with percent sequence identity across the respective evolutionary profiles (see Bioinformatics Results).

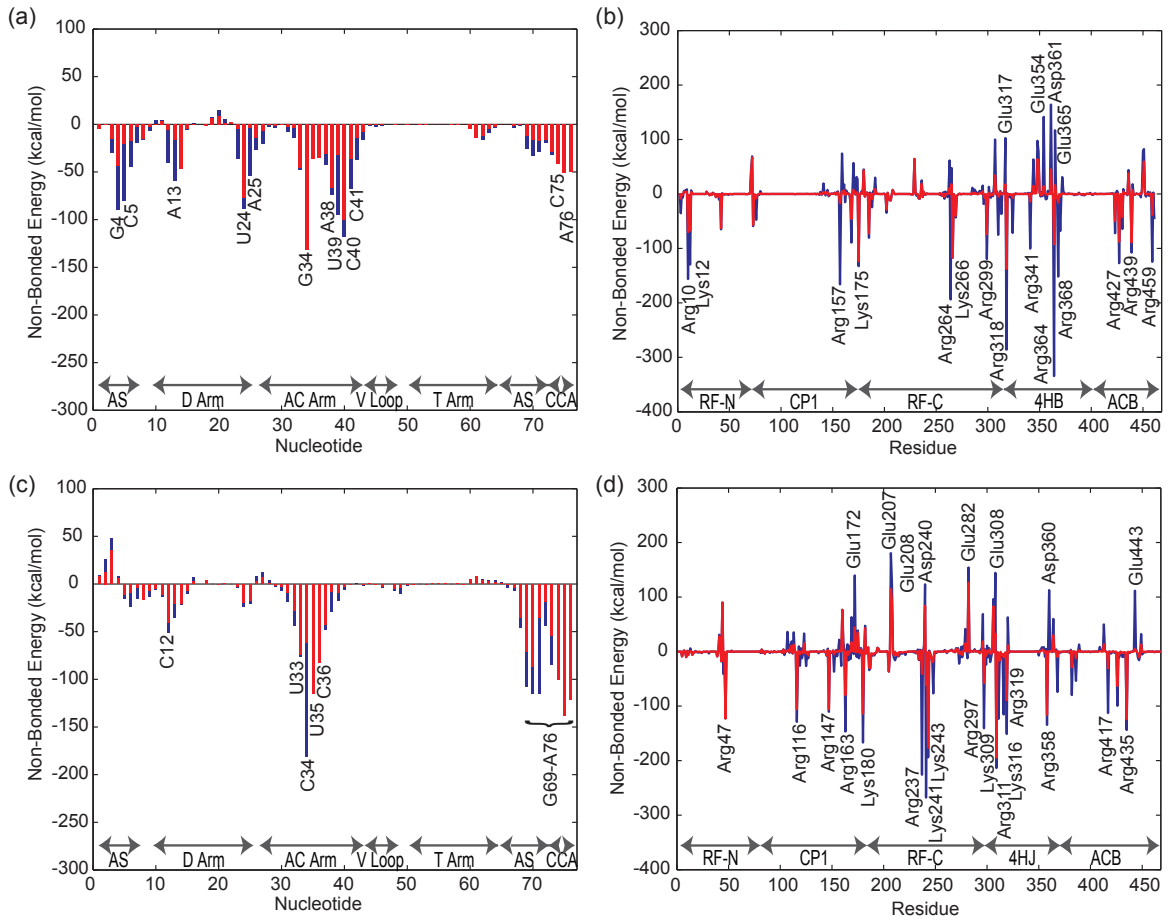


Figure 4.5: Interaction Energies of (a) $tRNA^{Cys}$ with (b) CysRS and (c) $tRNA^{Glu}$ and (d) GluRS. labeled nucleotides and residues are ±50 and ±100 kcal/mol respectively. The interaction energy is shown in blue while the energy masked by conservation is shown in red. Energies are averaged over the last 16 ns of each 20 ns simulation in the Pre-transfer state.

Identity elements energetically stabilize the aaRS·tRNA binding interface To validate the changed interface in the CysRS·$tRNA^{Cys}$ model, an examination of the identity elements in $tRNA^{Cys}$ making contacts with conserved residues at the interface was conducted and compared with the productively bound GluRS·$tRNA^{Glu}$. The identity elements are for $tRNA^{Cys}$ and $tRNA^{Glu}$ [65, 67] are listed below with their percent

sequence identity over the tRNA evolutionary profiles. The identity elements in $tRNA^{Cys}$ are G2·C71 (59/59%), C3·G70 (51/49%), G34 (100%), C35 (100%), A36 (100%), G15·G48 (87/5%), A13·A22 (28/85%), and U73 (90%) [65]. Of these, G34 in the anticodon and A13 in the core form strongly energetic contacts with CysRS. G1·C72 (90/49%), U2·A71 (61/29%), C4·G69 (68/68%), U34 (29%), U35 (100%), A37 (100%), U11·A24 (94/90%), U13·G22·A46 (65/58/71%) and the absence of a nucleotide at position 47 are the identity elements for $tRNA^{Glu}$. Of these, C72, A71, G69 in the acceptor stem, U34 and U35 in the anticodon loop make significant energetic contributions to the binding of the interface. These results would predict that $tRNA^{Cys}$ is not bound with the same strength or specificity as $tRNA^{Glu}$ based on identity elements alone. The $tRNA^{Cys}$ elements in the acceptor stem that dock in the CysRS active could be predicted to bind less tightly because $tRNA^{Cys}$ has yet to be crystallized with the acceptor stem/CCA hairpin bound productively. This is in contrast with the $tRNA^{Glu}$, which is known from several crystal structures to have the acceptor stem/CCA hairpin bound tightly in the active site. This is presumably due to the GluRS requiring the presence of $tRNA^{Glu}$ during adenylate formation, while CysRS can form Cys-AMP in the absence of $tRNA^{Cys}$. The binding of the $tRNA^{Cys}$ core has been studied extensively and experiment has shown that the shape of the tRNA is crucial for recognition [190, 398, 196, 399]. This is consistent with $tRNA^{Cys}$ making strong core contacts through the identity element A13 as well as several other nucleotides in the D arm. However, the lack of strong contacts near the anticodon for $tRNA^{Cys}$ is inconsistent with both comparison with $tRNA^{Glu}$ and experiments on $tRNA^{Cys}$ [155, 192, 400]. G34 has been shown to be the most strongly bound nucleotide in the anticodon, as it is in Figure 4.5a, but C35 and A36 are also important for binding. Likewise in $tRNA^{Glu}$, C34, U35, and C36 are shown in Figure 4.5c to be make strong interactions, from a tightly bound wobble base (C34) to the slightly more loosely bound C36. The lack of strong interactions between the anticodon in $tRNA^{Cys}$ and CysRS is expected to be a modeling artifact. During the application of forces in IMD on the tRNA, the anticodon was pulled away from the CysRS anticodon binding domain. The ACB was

shown in a previous crystal structure [344] to be highly mobile in the absence of $tRNA^{Cys}$ and had high β–factors in the current structure [8] because it is linked by a single loop to the rest of the CysRS. During the IMD modeling, the ACB also changed its conformation, bending at the loop to continue binding the anticodon. However, the intricate network of hydrogen bonds between the anticodon bases and ACB residues was strained and weakened, breaking during the subsequent equilibration.

CysRS·$tRNA^{Cys}$ and GluRS·$tRNA^{Glu}$ show similar patterns of interface binding The highly electronegative tRNA phosphate backbone forms two kinds of interactions with the aaRS, attractive contacts with positively charged residues and repulsive contacts with negatively charged residues. Figures 4.5b and d show that the number of strongly attractive interactions with positively charged amino acids is larger by nearly a factor of two than the repulsive contacts with negatively charged residues. Many of strong attractive interactions are balanced by repulsive ones that are close in sequence (e.g. repulsive Glu317 counteracts strongly attractive Arg318 that binds Ψ39 and C40 phosphate backbone in the anticodon stem). Interestingly, most of the strongly attractive contacts are made with CysRS residues that are conserved across all domains of life (46% vs. 37% over the entire enzyme). The exceptions are Arg157 and Arg368, which are in the CP1 insertion and 4HB respectively, the two most divergent regions of the enzyme. However, none of the repulsive interactions are strongly conserved because they are all in the 4HB. This has similarities to the GluRS shown in Figure 4.5 (see Chapter 3 for the corresponding analysis).

The pattern of conserved residues has some similarity between the two aaRSs. Arg10 in the N-terminal half of the CysRS Rossman fold is structurally conserved as Arg163 in the CP1 insertion in GluRS. Arg264 and Lys266 in the C-terminal half of the CysRS Rossman fold are structurally similar to Lys241 and Lys243 in the GluRS RF-C. Arg299 in CysRS is near Arg297 in GluRS. The 4HB contacts in CysRS differ greatly from the 4HJ in GluRS because the 4HJ does not directly bind $tRNA^{Glu}$ while the 4HB binds the anticodon stem of $tRNA^{Cys}$. Finally, a pattern of two attractive residues bind the anticodon nucleotides in both systems. In CysRS, these are the highly conserved Arg427 and Arg439 which form

hydrogen bonds with G34 and C35 in the anticodon. In GluRS, these are Arg417 and Arg435 that perform the same function with C34 and U35. This hints at a structurally conserved interface across monomeric Class I aaRSs. The interface can be predicted to differ based on editing domains and size of anticodon binding domain, but will show similar patterns of conserved interfacial interactions. This also indicates that while the modeling artifacts distort the energetic interactions in $tRNA^{Cys}$, the CysRS interactions are consistent with those of a closely related aaRS. The position of nucleotides/residues in the modeled structure is shown to be accurate to a first approximation, but the exact interactions at the interface (i.e. hydrogen bonds or salt bridges) are not preserved.

4.4.5 Free energies of binding indicate the model is unstable

Similar to Chapter 3, the technique of Molecular Mechanics–Poisson-Boltzmann Surface Area (MM-PBSA) is used to calculate the average free energy of complex formation [366, 367], $\langle \Delta G_{binding} \rangle = \langle G_{XY} \rangle - (\langle G_X \rangle + \langle G_Y \rangle)$ where XY denotes a macromolecular complex formed from components X and Y. The free energy difference is expressed as a sum of terms that can be calculated from MD trajectories: $\langle \Delta G_{binding} \rangle = \langle \Delta E_{VdW} + \Delta E_{elec} + \Delta G_{polar} + \Delta G_{nonpolar} \rangle - T\langle \Delta S \rangle$. The time averages (indicated by $\langle \cdot \rangle$) are calculated over a selected number of frames from the 20 ns trajectory. The first two terms are the van der Waals and electrostatic interaction energies between the two molecules and are derived from molecular mechanics. G_{polar}, the polar solvation energy, is the energy required to move the complex and individual components from a dielectric of $\epsilon_{in} = 1.00$ to $\epsilon_{out} = 78.54$. The polar solvation energy is efficiently calculated using a Poisson-Boltzmann implicit continuum solvent model [372]. The nonpolar solvation energy, $G_{nonpolar}$, is the energy required to create a cavity in the solvent for a given system and is proportional to the solvent accessible surface area. The last term refers to the entropy changes on complex formation and was calculated from the molecular dynamics simulations using Schlitter's formula [373]. Since the free energy is calculated using the complex and system components from the same trajectory, contributions from

conformational changes in the CysRS or $tRNA^{Cys}$ are neglected in this analysis. Also, MM-PBSA is not generally successful at reproducing absolute binding free energy values accurately; the focus is therefore on stability of the CysRS·$tRNA^{Cys}$ relative to the GluRS·$tRNA^{Glu}$ in Chapter 3 and changes in free energy across the 20 ns trajectory.

	Pre-transfer CysRS	Pre-transfer GluRS
$\langle \Delta E_{VdW} \rangle$	-189.57	-244.49
	(0.34)	(0.34)
$\langle \Delta E_{coulomb} + \Delta G_{polar} \rangle$	213.68	-742.21
	(0.97)	(6.42)
$\langle \Delta G_{nonpolar} \rangle$	-25.18	-29.07
	(0.04)	(0.02)
$-T\Delta S$	40.40	36.20
$\langle \Delta G_{binding} \rangle$	39.33	-8.79
	(0.80)	(0.65)

Table 4.2: MM-PBSA free energy differences for tRNA binding in kcal/mol for the Pre-transfer states containing CysRS and GluRS (from Chapter 3). The 95 % confidence interval for each value is ± the number shown below in parentheses. Standard deviations for the $\langle \Delta G_{binding} \rangle$ were 17-24 kcal/mol. The dielectric was set to 1.0 in each run.

In comparing the pre-transfer state free energies over the last 5 ns (see Table 4.2, the binding of $tRNA^{Cys}$ to CysRS in the modeled structure differs from the GluRS system by 50 kcal/mol. This significant decrease in binding affinity in the CysRS·$tRNA^{Cys}$ results from both of the modeling artifacts seen in the local energetics analysis: the anticodon loop weakly interacting with the anticodon binding domain and the weak interactions between the acceptor stem and catalytic domain. Further, the free energy over time varies more in the CysRS pre-transfer state (range of 30 kcal/mol) than in the partially undocked GluRS Post (no AMP/$GluNH_2$) state (range of 20 kcal/mol) over 20 ns (see Figure 4.6. While it is possible that CysRS binds $tRNA^{Cys}$ less tightly than GluRS binds $tRNA^{Glu}$ due to presence of modified nucleotides in $tRNA^{Cys}$, it is questionable as to whether the free energy of complex binding will change to this degree in the pre-transfer state. These results indicate that while the CysRS·$tRNA^{Cys}$ model was shown to be well modeled on a coarse grained level by the Class I aaRS structure alignment and active site interactions, the finer details of the model may be lacking. It is possible that several hydrogen bonds and salt bridges are missing in the model, interactions that would stabilize the complex interface

during the simulation. Future modeling attempts should focus on optimizing hydrogen bonds and salt bridges.

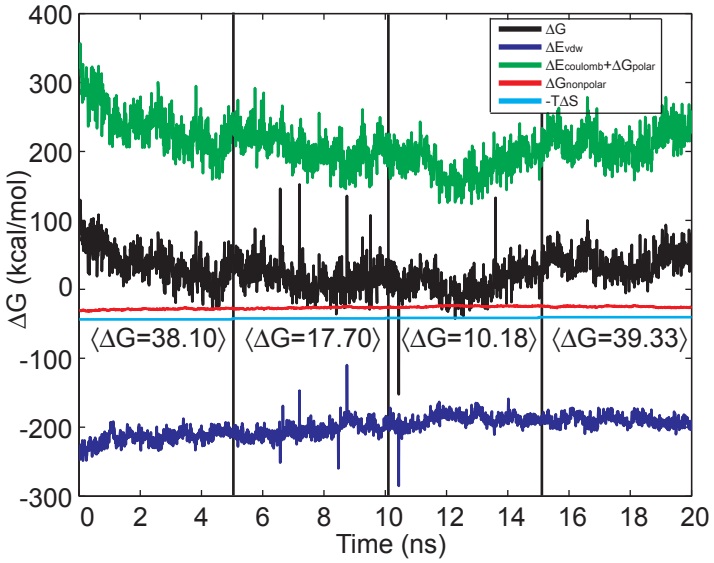


Figure 4.6: $\Delta G(t)$ averaged over 5 ns windows from the 20 ns Pre-transfer simulation for uncharged $tRNA^{Cys}$ binding CysRS. The components of the total $\Delta G(t)$ sampled every 40 ps are given over the 20 ns simulation. The entropy averaged over 5 ns windows is included in $\Delta G(t)$ as well as $\langle \Delta G \rangle$ computed over 5 ns windows.

4.5 Conclusion

The unproductive CysRS·$tRNA^{Cys}$ crystal structure 1U0B is considered a trivial modeling problem by aaRS experts, but our calculations indicate that the conformation of the bound tRNA is significantly different than any other aaRS·tRNA structure. Alignment of available Class I aaRS·tRNA complex structures shows that $tRNA^{Cys}$ is incorrectly bound at the core and along the acceptor stem in addition to A76 on the CCA hairpin being docked in the AMP binding site. Using several modeling techniques, we obtained a productive complex that is valid on a coarse grain level in comparison to the Class I structure alignment and active site patterns of substrate binding. Local energetics and free energy analyses show that while there are several highly conserved residue·nucleotide interactions at the interface that are consistent with the bioinformatic analysis, the

acceptor stem and anticodon loop are less tightly bound than in other aaRS·tRNA complexes, allowing the complex to begin dissociating before aminoacylation occurs. Therefore, our CysRS·$tRNA^{Cys}$ model structure is considered sufficient to make general predictions, such as the identity of the conserved triad residues binding the α–ammonium group of the charging cysteine, but not accurate for performing more rigorous analyses. The modeling artifact introduced at the anticodon arm by applying forces on the tRNA to correct the crystallization artifact of the incorrectly docked tRNA core breaks essential hydrogen bonds. Estimation of the acceptor stem and CCA hairpin conformation based on other aaRS·tRNA structures yields a rough approximation of the correct orientation, but strong interactions between the acceptor stem/CCA hairpin and the protein do not form in the short 20 ns timespan of our molecular dynamics simulation. Therefore, future attempts to model this structure should include development of computational tools to optimize the model protein·tRNA interactions on an atomic level.

4.6 Acknowledgements

The author thanks ZLS group members, particularly John Eargle and Elijah Roberts, for many helpful conversations as well as Lennart Nilsson for pseudouridine parameters and Nathan Baker for APBS assistance. Funding and supercomputer time was provided by: NSF (MCB04-46227), NIH Molecular Biophysics Training Grant (PHS5T32GM08276), NIH Chemical Biology Training Grant (5T32GM070421), NCSA LRAC (MCA03T027), NIH (P41-RR05969), and NSF CRIF (0541659).

APPENDIX A

ENTROPY DIFFERENCE GRAPHS FOR ALL TRNAS BY DOMAIN/SPECIFICITY

This appendix shows the entropy difference graphs described in Chapter 2 for all tRNAs by domain of life/amino acid specificity. The difference graphs are shown with those positions known to be identity elements from previous studies labeled. Those with representative structures have the positions at the aaRS·tRNA interface colored green. The main tuning element, G51·C63, has been colored red in all plots to indicate whether it and surrounding nucleotides may be tuning elements in each domain/specificity. Each caption contains the references to the numerous studies reporting identity elements and the citation of the crystal structures used to locate the nucleotides at the interface. The PDB codes for these structures are included on the relevant graphs.

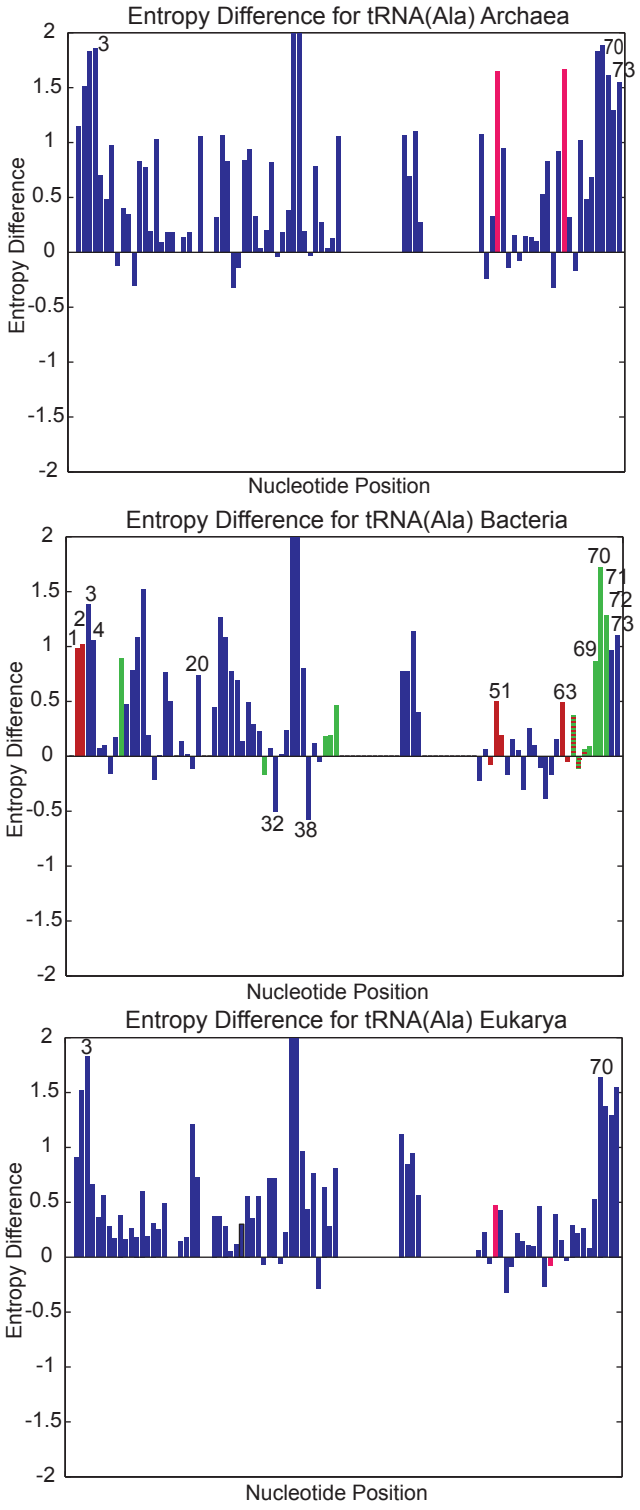


Figure A.1: Identity element references for $tRNA^{Ala}$: [119, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172] Nucleotides at the interface in the bacterial system were determined by partial digestion [153].

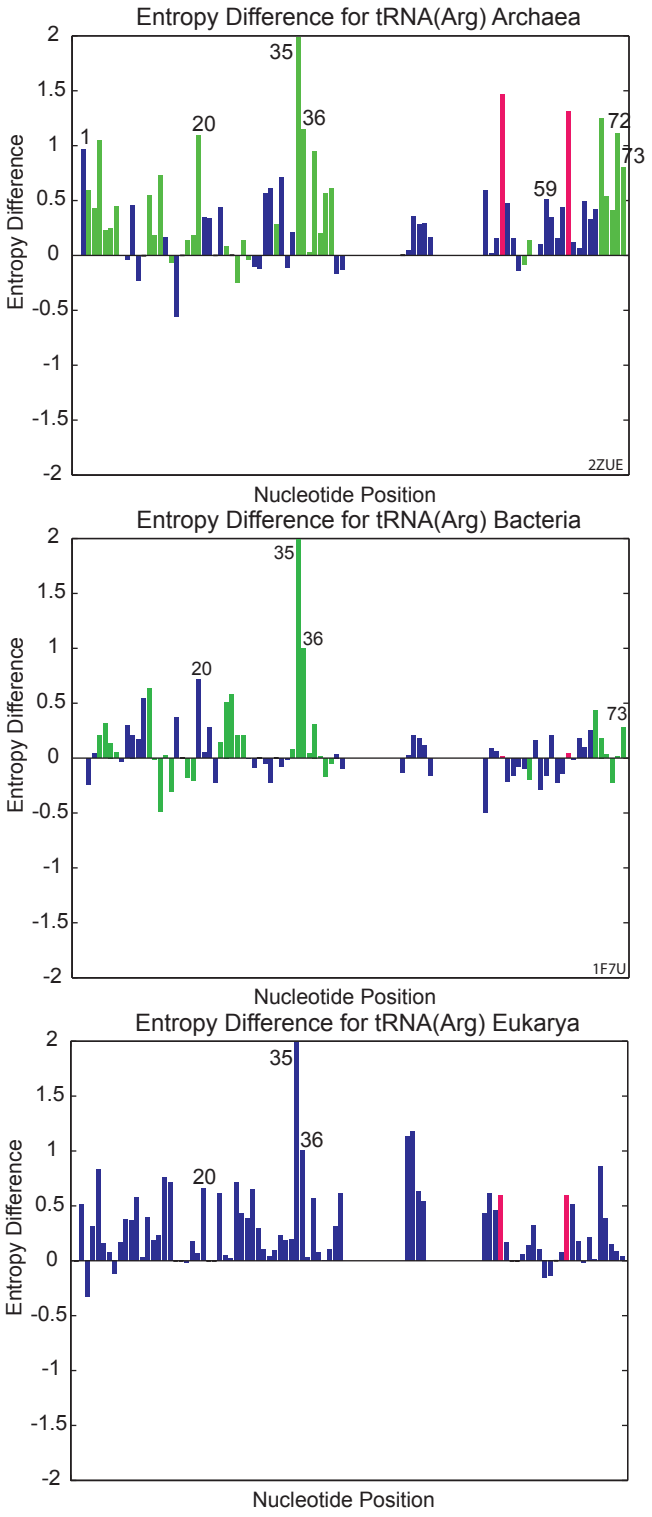


Figure A.2: Identity element references for $tRNA^{Arg}$:
[173, 174, 175, 176, 177, 5, 178, 179, 180, 172] Nucleotides at the interface in the archaeal and bacterial systems were determined from [401] and [343] respectively.

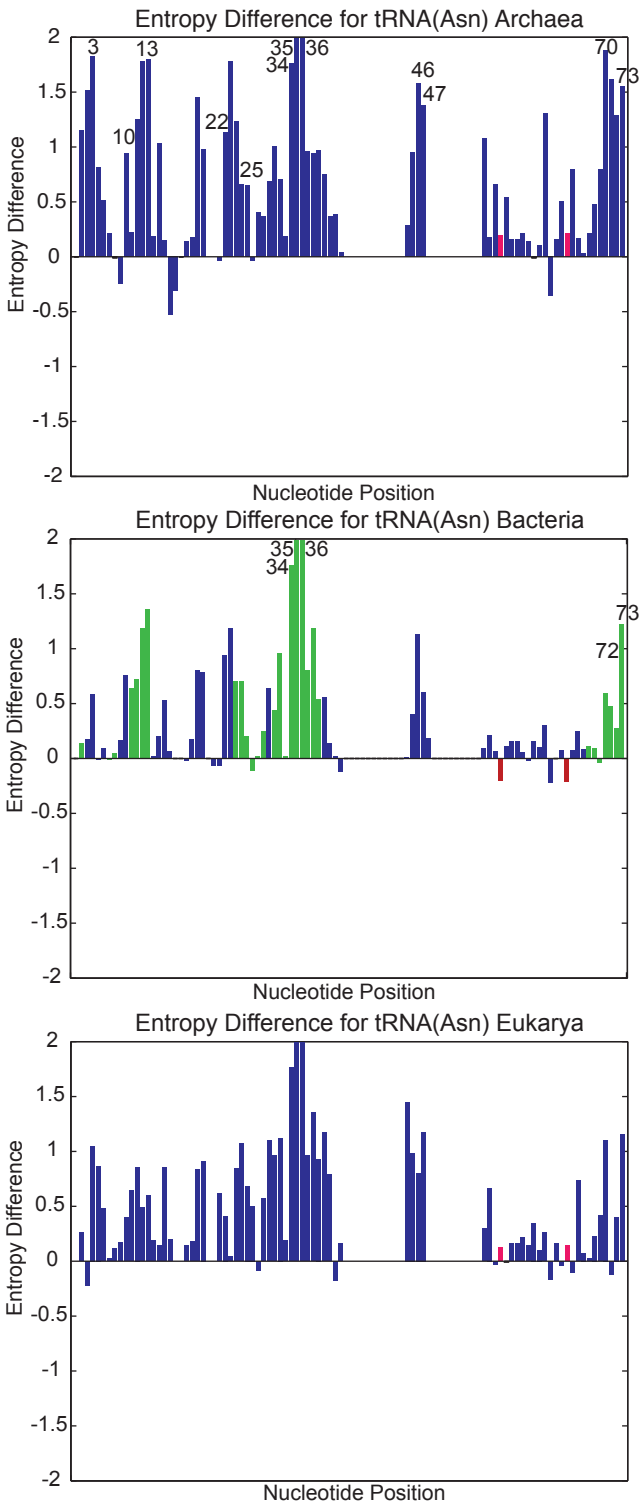


Figure A.3: Identity element references for $tRNA^{Asn}$: [189, 402, 254, 403, 172] Nucleotides at the interface in the bacterial system were determined by analogy to the homologous *E. coli* AspRS [404].

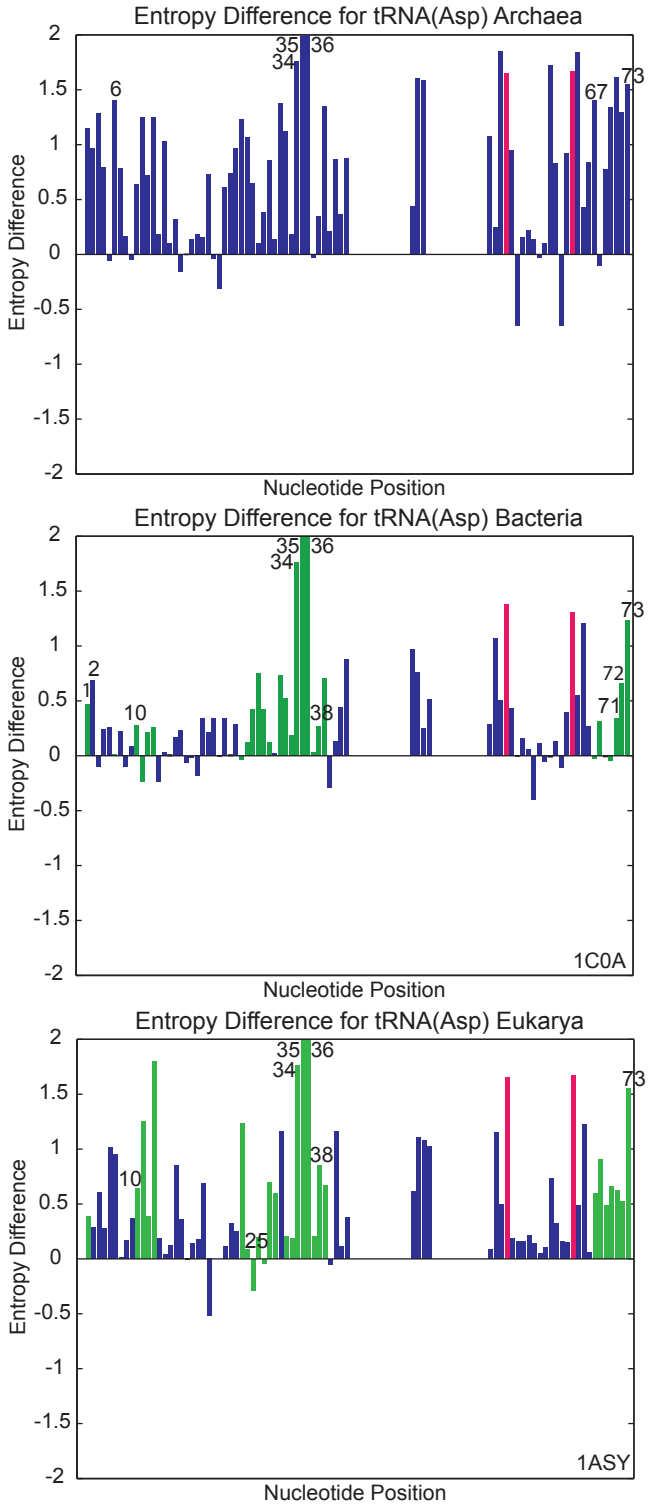


Figure A.4: Identity element references for $tRNA^{Asp}$:
[181, 182, 183, 184, 185, 186, 187, 5, 188, 172] Nucleotides at the interface in the bacterial and eukaryal systems were determined from [404] and [405] respectively.

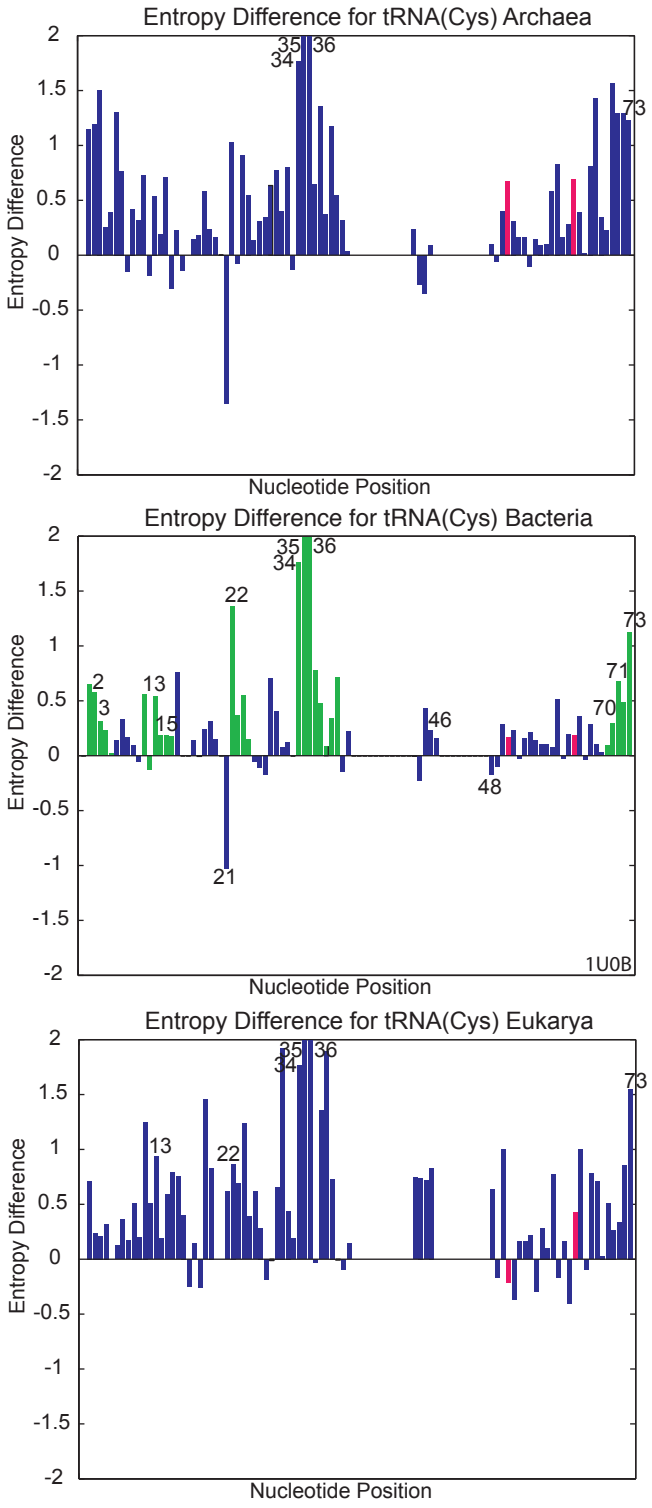


Figure A.5: Identity element references for $tRNA^{Cys}$:
[155, 189, 190, 191, 192, 193, 194, 195, 196, 197, 5, 198, 199, 172] Nucleotides at the interface in the bacterial system were determined from the modeled *E. coli* structure [8] and Chapter 4.

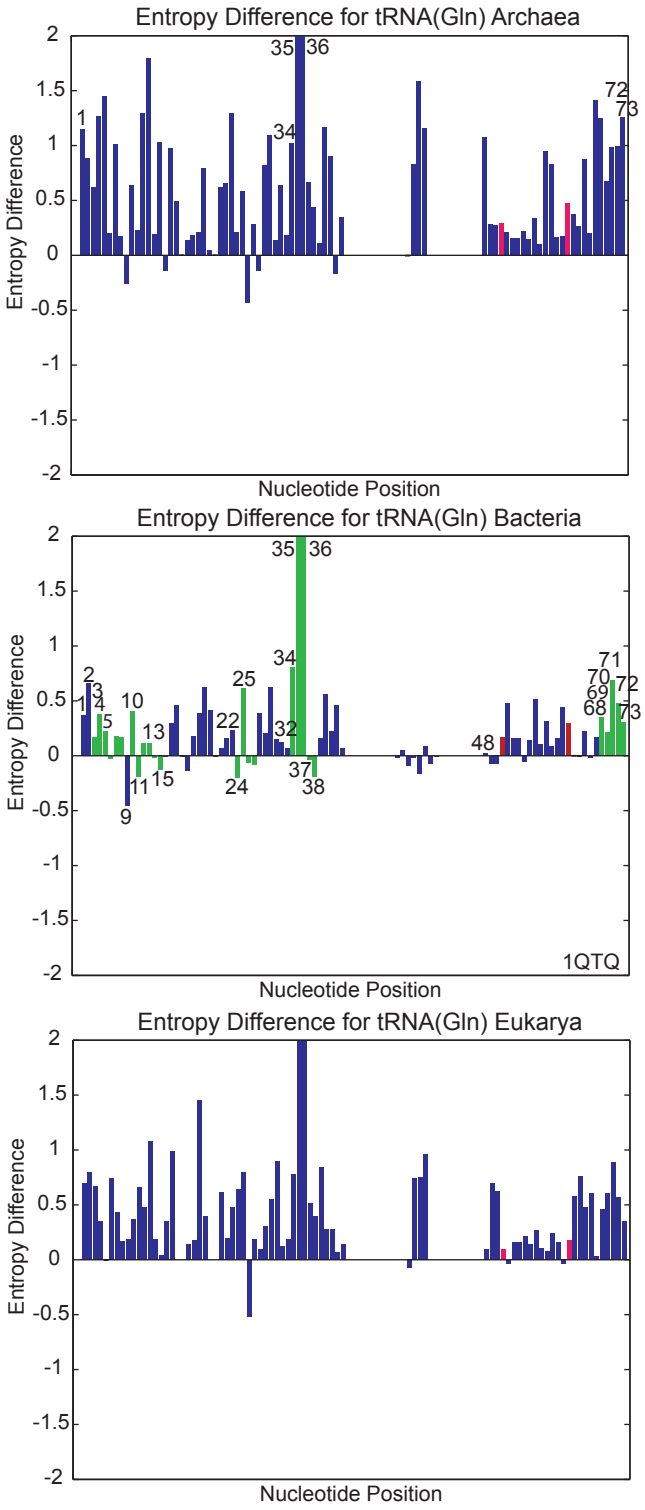


Figure A.6: Identity element references for $tRNA^{Gln}$:
[200, 201, 202, 203, 123, 100, 154, 204, 148, 172] Nucleotides at the interface in the bacterial system were determined from [342].

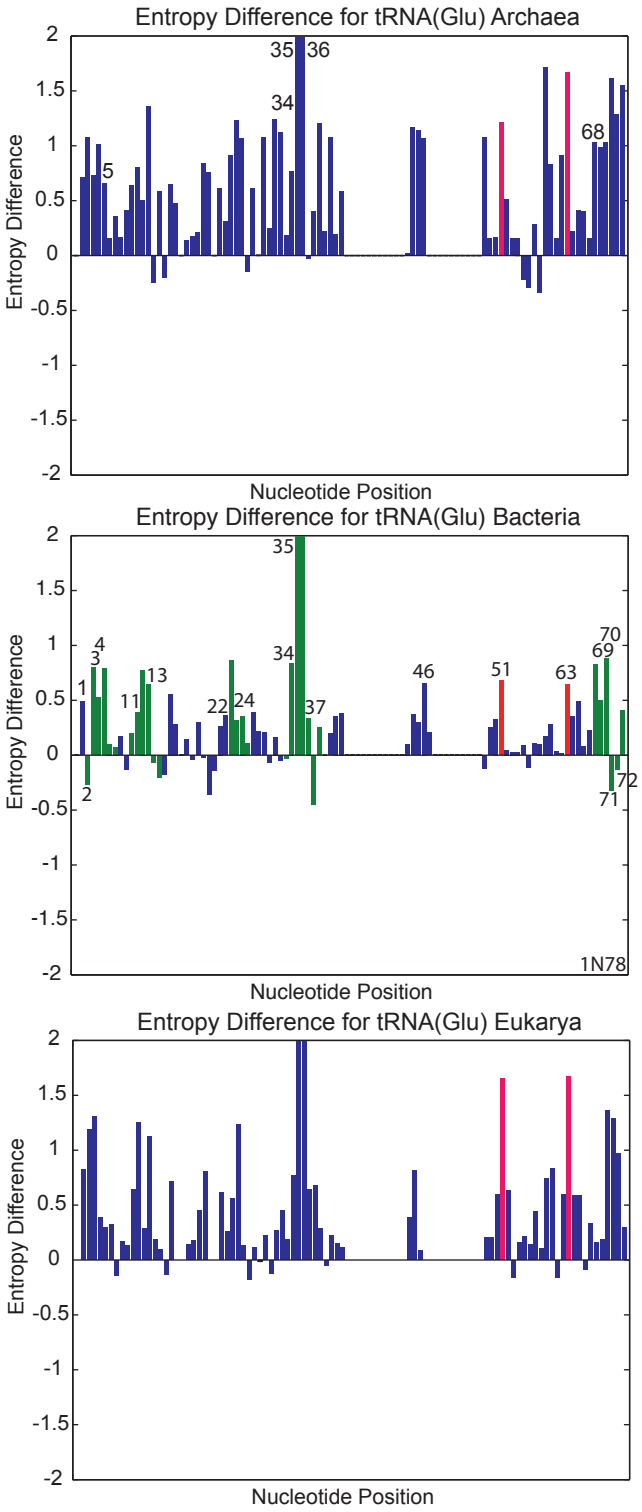


Figure A.7: Identity element references for $tRNA^{Glu}$: [205, 206, 207, 208, 209, 172] Nucleotides at the interface in the bacterial system were determined from [316].

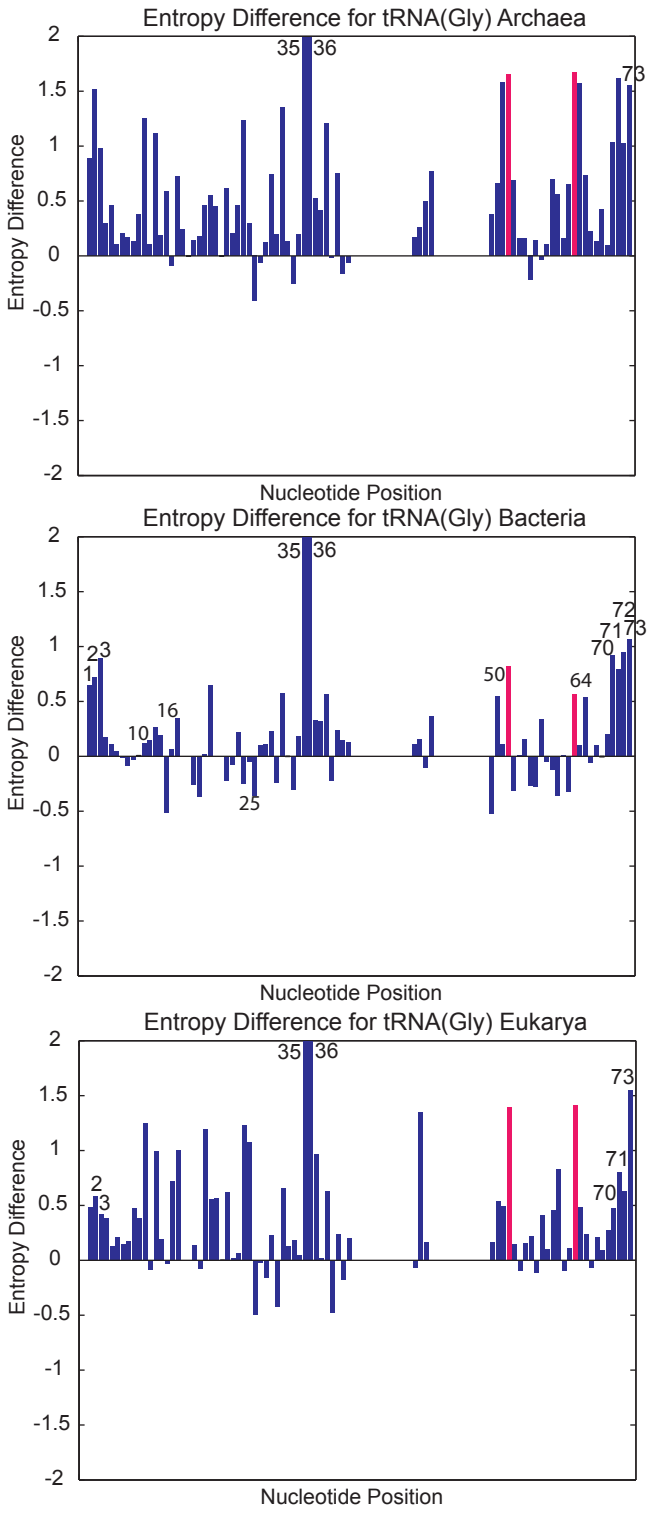


Figure A.8: Identity element references for $tRNA^{Gly}$:
[189, 210, 211, 212, 213, 214, 212, 215, 172]

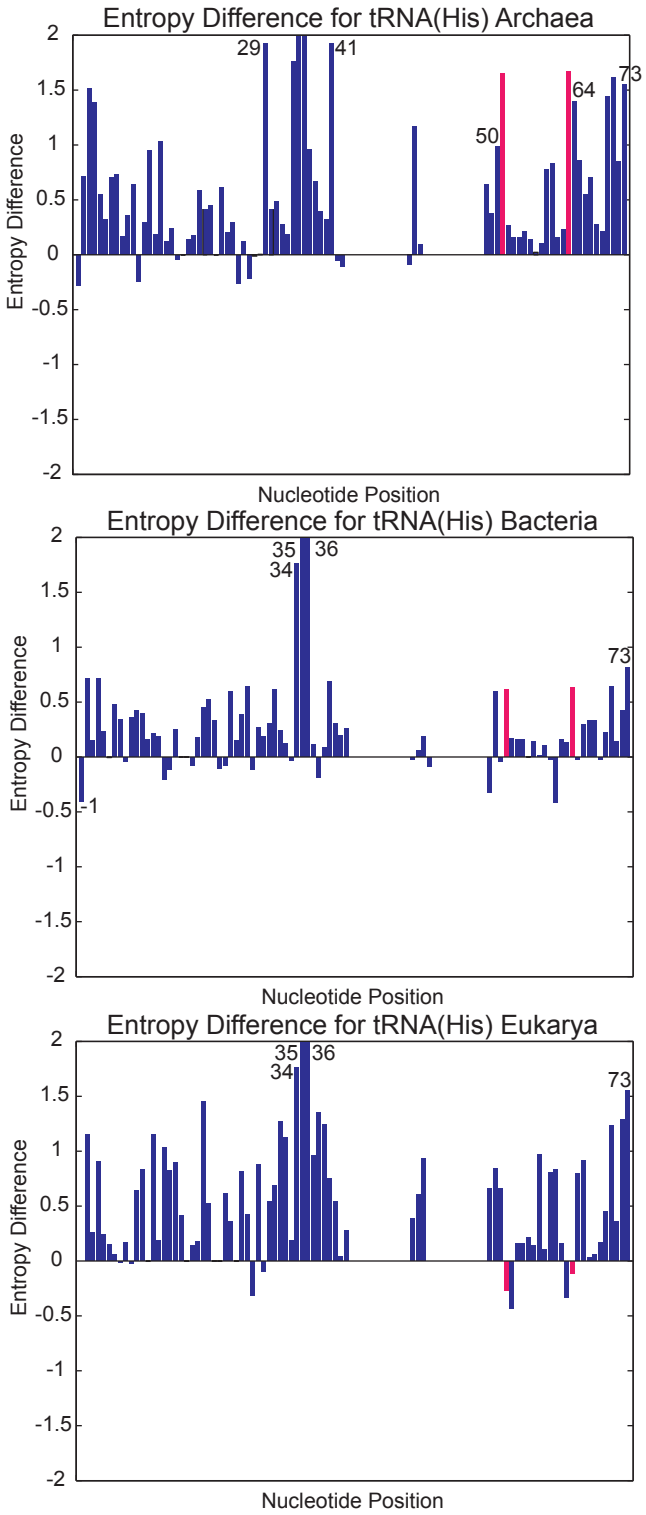


Figure A.9: Identity element references for $tRNA^{His}$:
[189, 211, 216, 217, 218, 219, 220, 221, 222, 223, 224, 172]

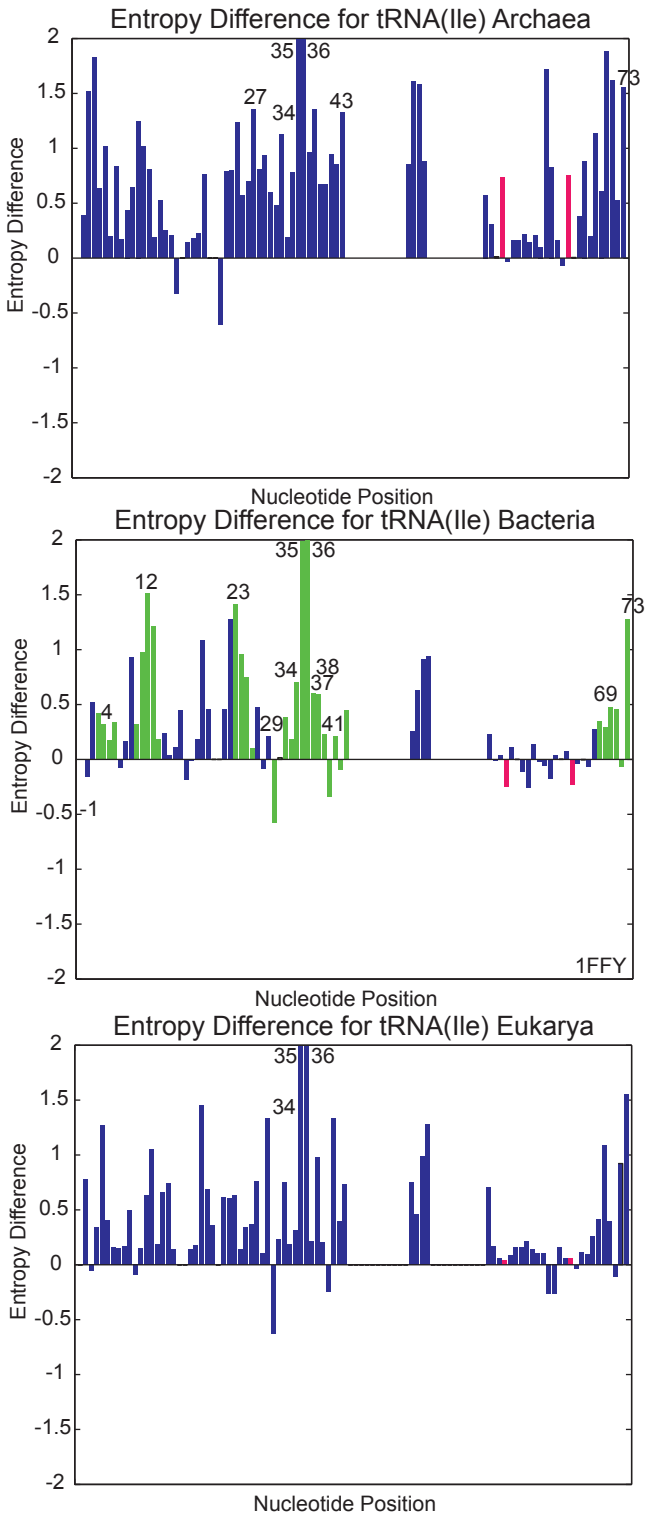


Figure A.10: Identity element references for $tRNA^{Ile}$: [225, 31, 226, 227, 172] Nucleotides at the interface in the bacterial system were determined from [345].

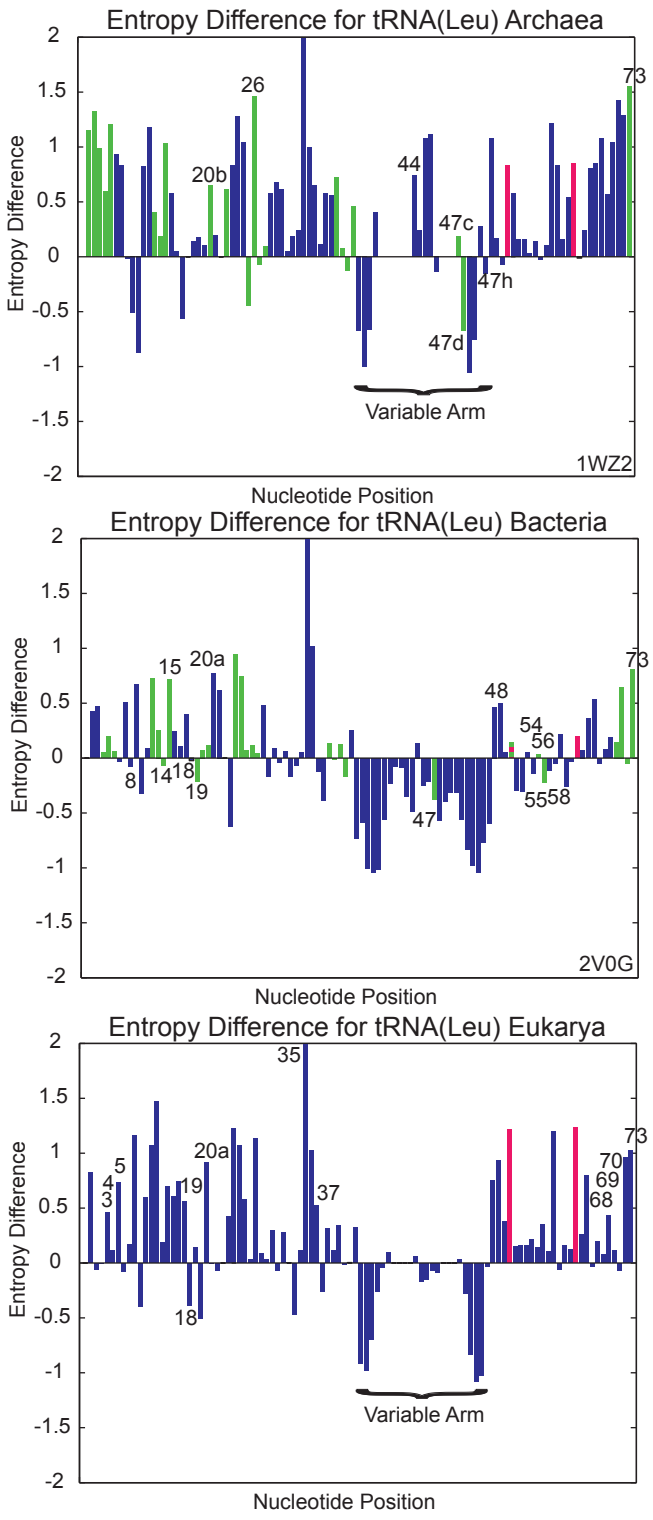


Figure A.11: Identity element references for $tRNA^{Leu}$:
[228, 229, 230, 231, 232, 233, 234, 235, 5, 236, 237, 172] Nucleotides at the interface in the archaeal and bacterial systems were determined from [401] and [343] respectively.

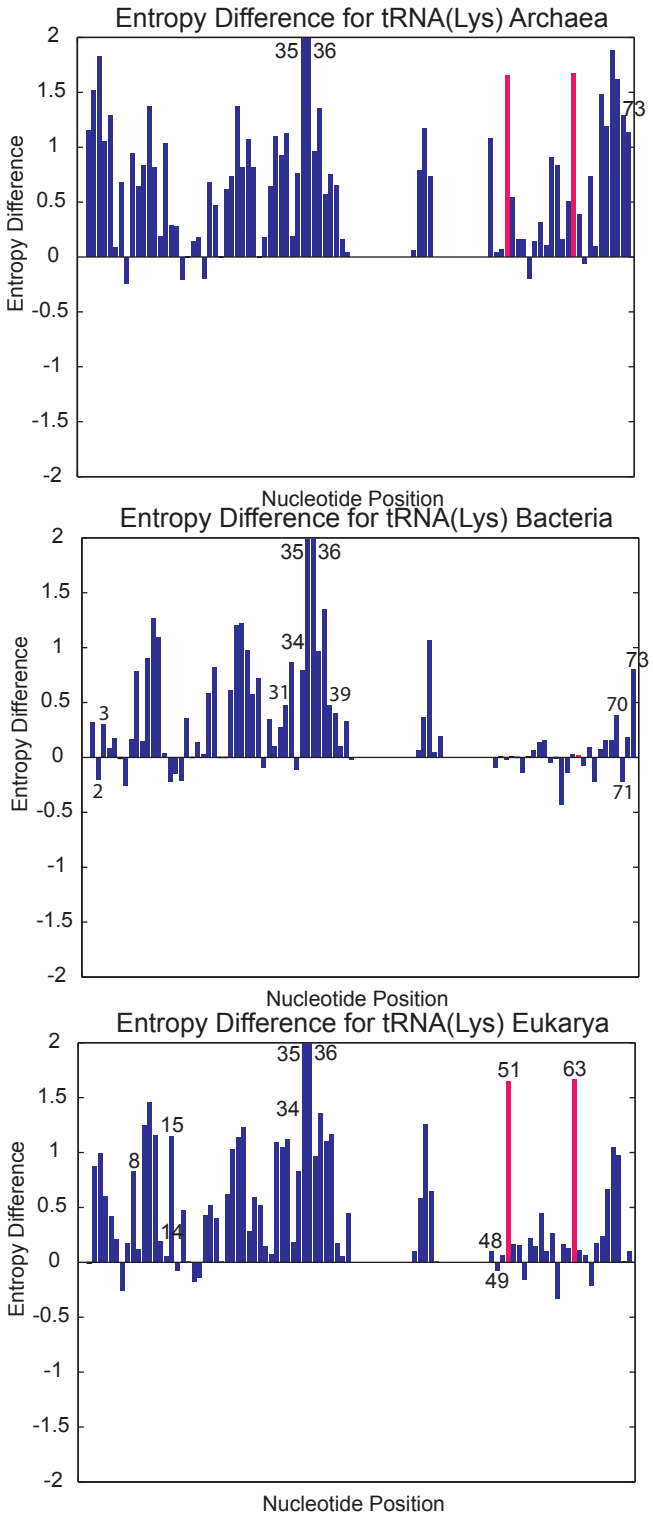


Figure A.12: Identity element references for $tRNA^{Lys}$:
[205, 175, 176, 238, 239, 240, 241, 242, 243, 244, 245, 246, 172]

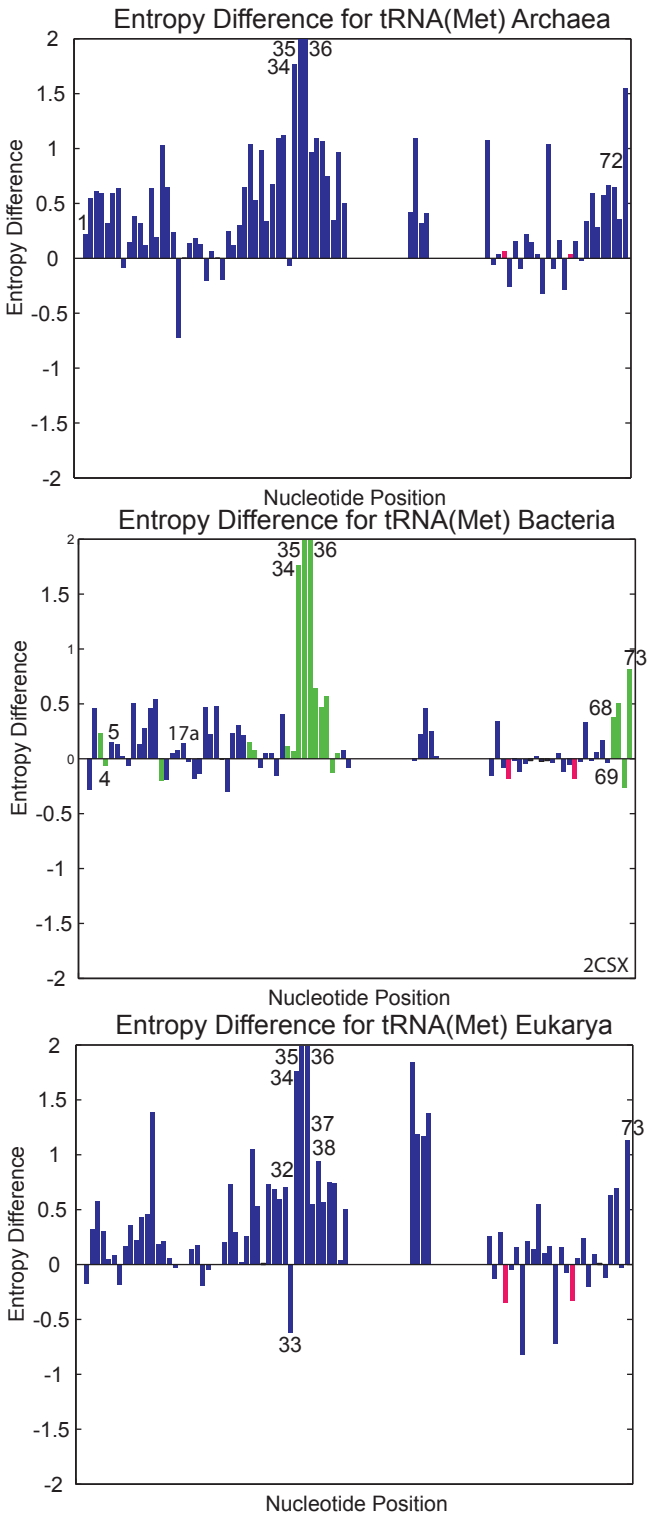


Figure A.13: Identity element references for $tRNA^{Met}$:
[247, 248, 249, 250, 251, 252, 253, 254, 172] Nucleotides at the interface in the bacterial system were determined from [406].

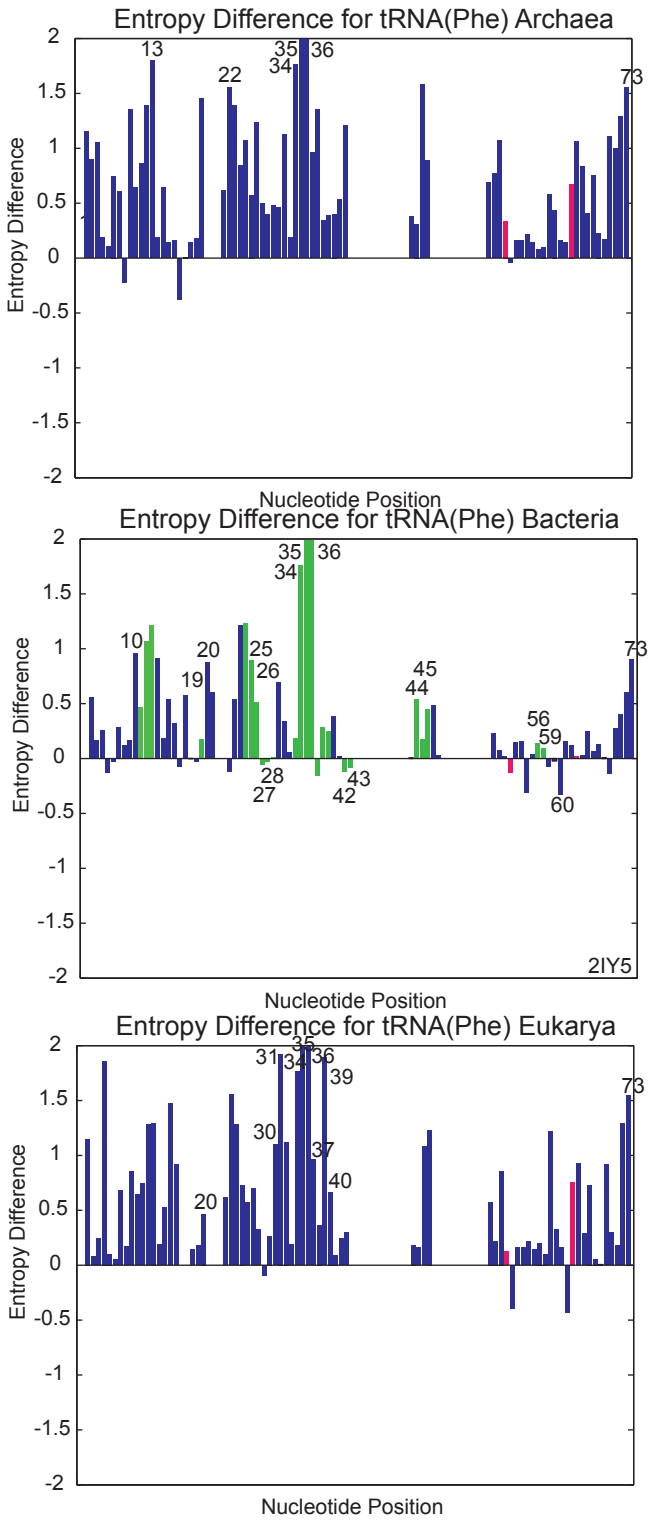


Figure A.14: Identity element references for $tRNA^{Phe}$:
[225, 255, 256, 257, 258, 259, 260, 161, 261, 172] Nucleotides at the interface in the bacterial system were determined from [407].

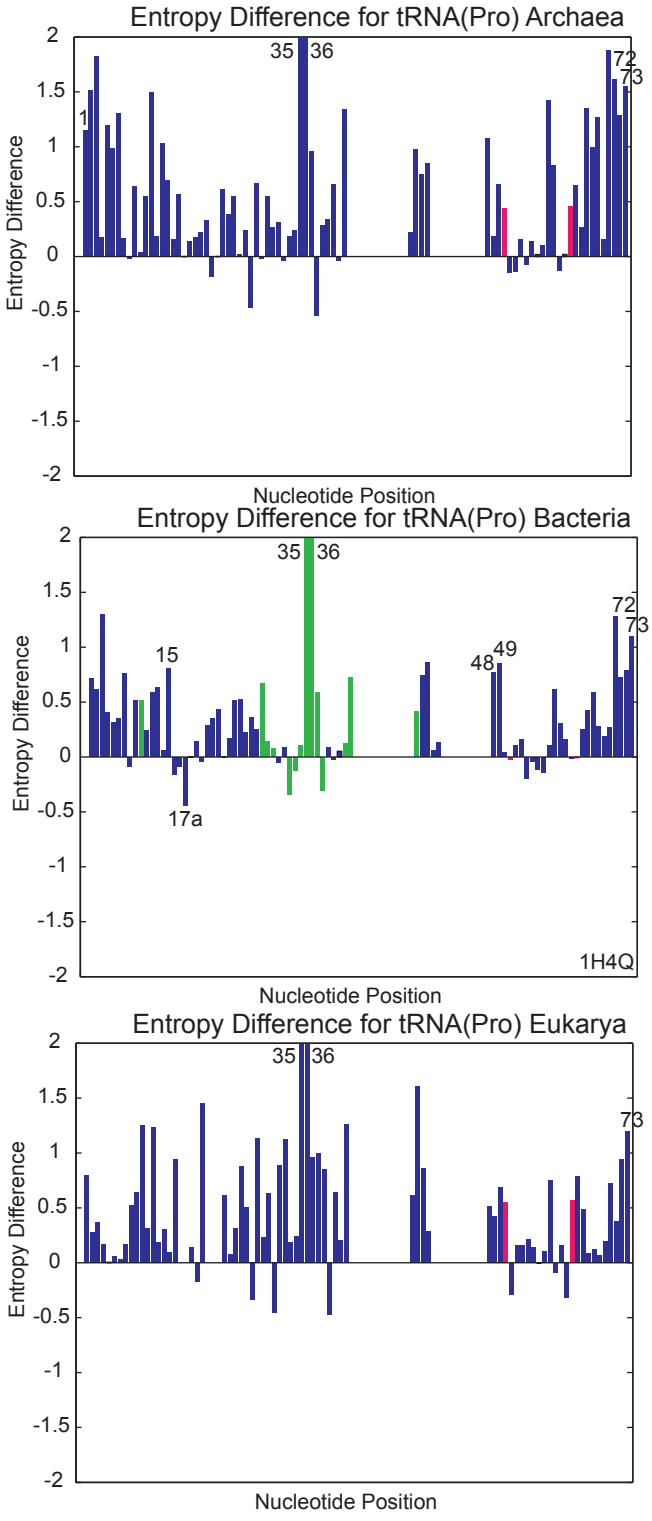


Figure A.15: Identity element references for $tRNA^{Pro}$:
[189, 262, 263, 264, 265, 266, 267, 254, 172] Nucleotides at the interface in the bacterial system were determined from [408].

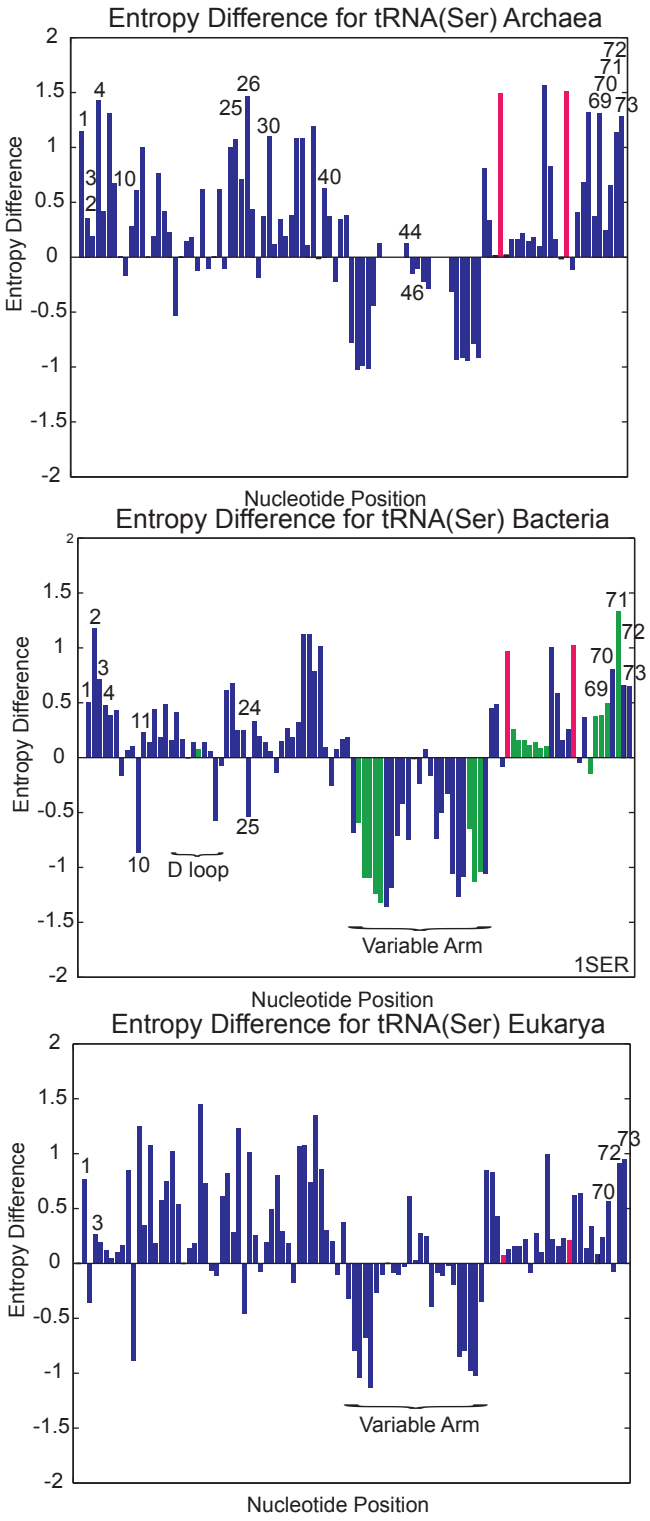


Figure A.16: Identity element references for $tRNA^{Ser}$: [228, 231, 268, 200, 269, 270, 271, 272, 273, 274, 5, 275, 276, 277, 254, 278, 230, 279, 280, 281, 172] Nucleotides at the interface in the bacterial system were determined from [278].

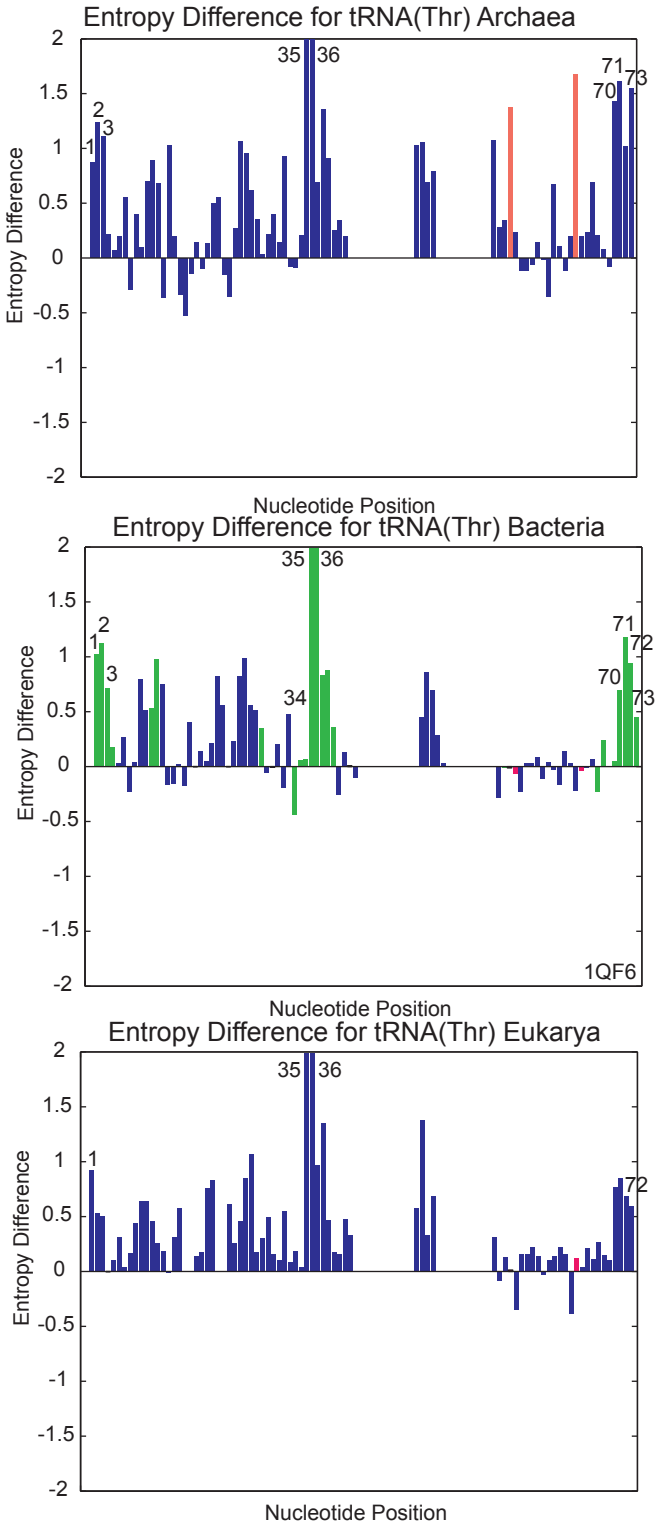


Figure A.17: Identity element references for $tRNA^{Thr}$: [282, 283, 284, 285, 286, 287, 172] Nucleotides at the interface in the bacterial system were determined from [397].

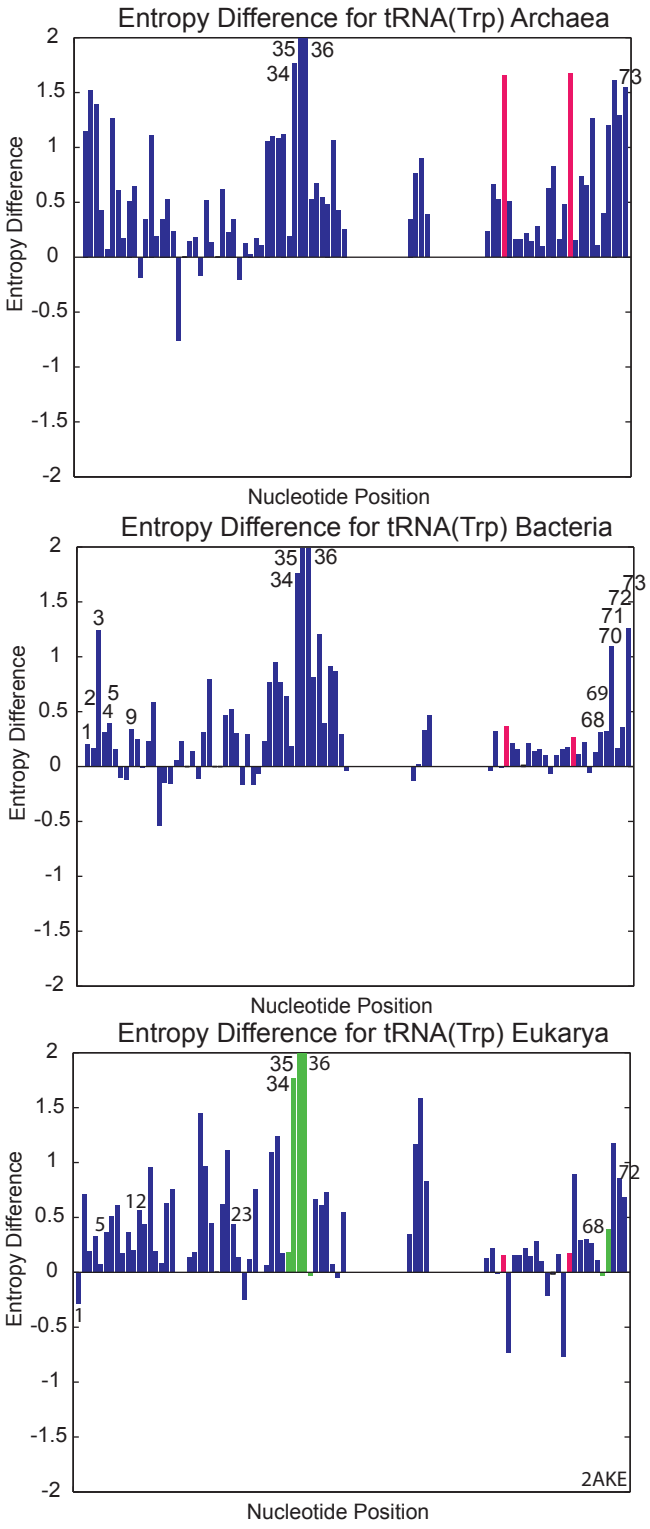


Figure A.18: Identity element references for $tRNA^{Trp}$:
[288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 172] Nucleotides at the interface in the eukaryal system were determined from [298].

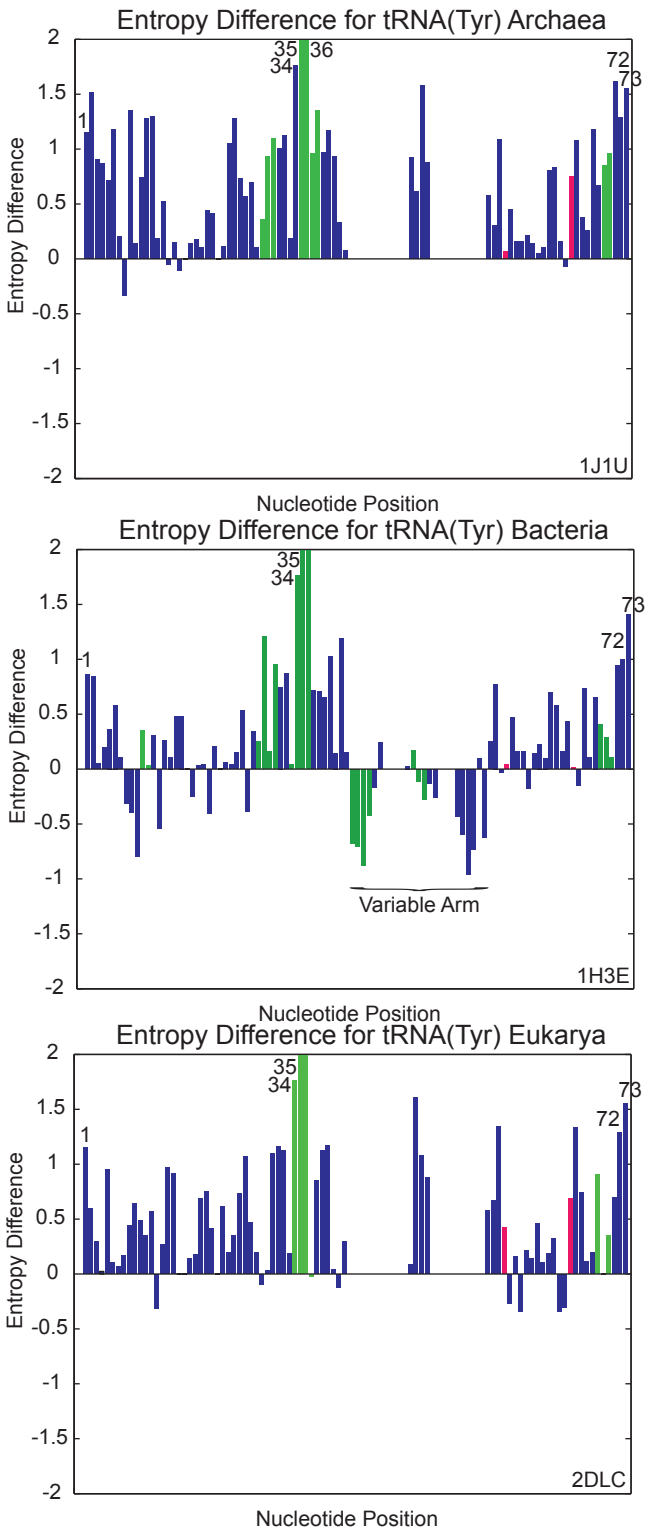


Figure A.19: Identity element references for $tRNA^{Tyr}$:
[161, 300, 268, 301, 302, 249, 303, 304, 305, 306, 307, 308, 60, 172] Nucleotides at the interface in the archaeal, bacterial, and eukaryal system were determined from [60], [349], and [381] respectively.

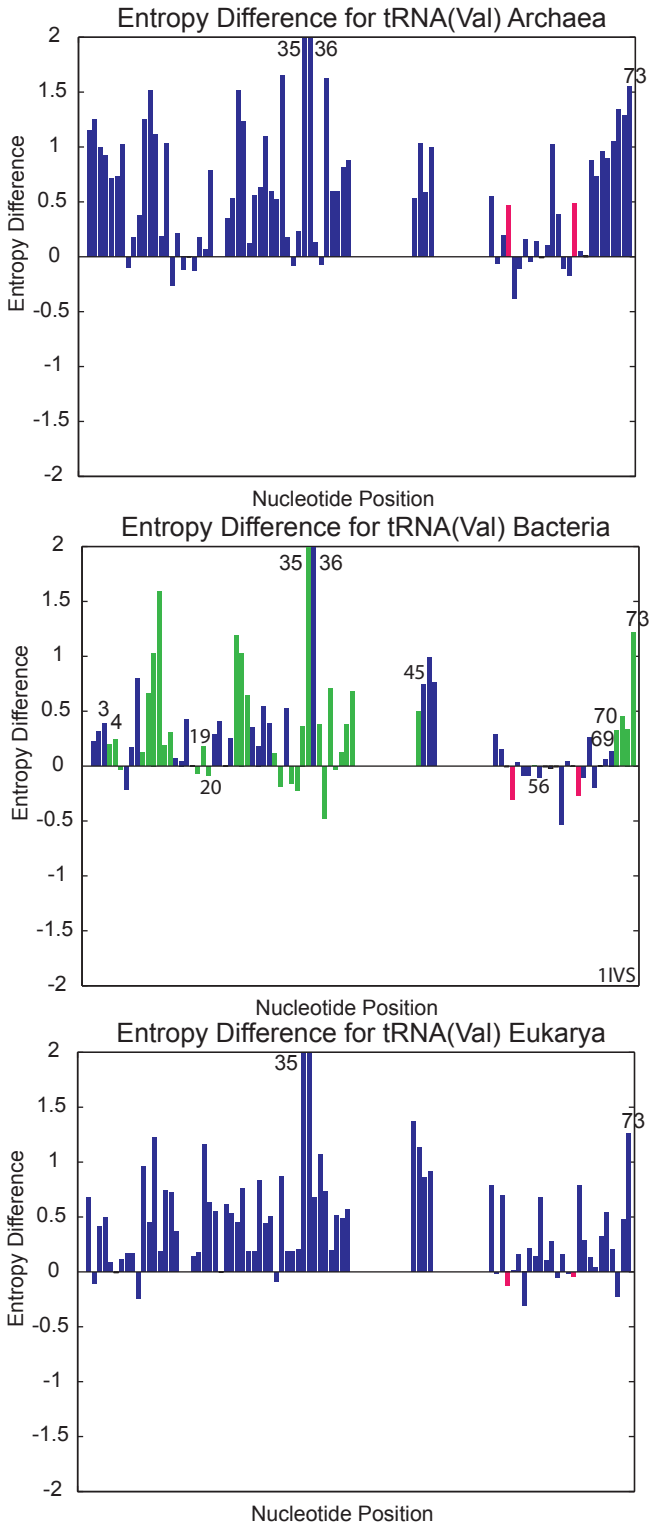


Figure A.20: Identity element references for $tRNA^{Val}$:
[309, 225, 310, 311, 312, 313, 5, 314, 172] Nucleotides at the interface in the bacterial system were determined from [348].

REFERENCES

[1] Ibba, M (2010) Transfer RNA. *FEBS Lett.* 584:251.

[2] Crick, FH (1968) The origin of the genetic code. *J. Mol. Biol.* 38:367–379.

[3] Hopper, A, Pai, D, Engelke, D (2010) Cellular dynamics of tRNAs and their genes. *FEBS Lett.* 584:310–317.

[4] Putz, J, Giege, R, Florentz, C (2010) Diversity and similarity in the tRNA world: Overall view and case study on malaria-related tRNAs. *FEBS Lett.* 584:350–358.

[5] (2005) *The Aminoacyl-tRNA synthetases* (Landes Bioscience, Georgetown, Texas).

[6] Sprinzl, M, Vassilenko, KS (2005) Compilation of tRNA sequences and sequences of tRNA genes. *Nucl. Acids Res.* 33:139–140.

[7] Grosjean, H, de Crecy-Lagard, V, Marck, C (2010) Deciphering synonymous codons in the three domains of life: Co-evolution with specific tRNA modification enzymes. *FEBS Lett.* 584:252–264.

[8] Hauenstein, S, Zhang, CM, Hou, YM, Perona, JJ (2004) Shape-selective RNA recognition by cysteinyl-tRNA synthetase. *Nat. Struct. Mol. Biol.* 11:1134–1141.

[9] Kim, S, Suddath, F, Quigley, G, McPherson, A, Sussman, J, Wang, A, Seeman, N, Rich, A (1974) Three-dimensional tertiary structure of yeast phenylalanine transfer RNA. *Science* 185:435–440.

[10] Robertus, JD, Ladner, JE, Finch, JT, Rhodes, D, Brown, RS, Clark, BF, Klug, A (1974) Structure of yeast phenylalanine tRNA at 3 A resolution. *Nature* 250:546–551.

[11] Giege, R (2008) Toward a more complete view of tRNA biology. *Nat. Struct. Mol. Biol.* 15:1007–1014.

[12] Walker, S, Engelke, D (2006) Ribonuclease P: the evolution of an ancient RNA enzyme. *Crit. Rev. Biochem. Mol. Biol.* 41:77–102.

[13] Xiao, S, Fierke, C.A. Engelke, D (2002) Eukaryotic ribonuclease P: a plurality of ribonucleoprotein enzymes. *Annu. Rev. Biochem.* 71:165–189.

[14] Zuo, Y, Deutscher, M (2001) Exoribonuclease superfamilies: structural analysis and phylognetic distribution. *Nuc. Acids Res.* 29:1017–1026.

[15] Reyes, V, Abelson, J (1988) Substrate recognition and splice site determination in yeast tRNA splicing. *Cell* 55:719–730.

[16] Marck, C, Grosjean, H (2003) Identification of BHB splicing motifs in intron-containing tRNAs from 18 archaea: Evolutionary implications. *RNA* 9:1516–1531.

[17] Sugahara, J, Yachie, N, Arakawa, K, Tomita, M (2007) In silico screening of archaeal tRNA-encoding genes having multiple introns with bulge-helix-bulge splicing motifs. *RNA* 13:671–681.

[18] Sugahara, J, Kikuta, K, Fujishima, K, Yachie, N, Tomita, M, Knaai, A (2008) Comprehensive analysis of archaeal tRNA genes reveals rapid increase of tRNA introns in the order thermproteales. *Mol. Biol. Evol.* 25:2709–2716.

[19] Randau, L, Munch, R, Hohn, R, Jahn, M, Söll, D (2005) Nanoarchaeum equitans creates functional tRNAs from separate genes for the 5'= and 3'-halves. *Nature* 433:537–541.

[20] Randau, L, Pearson, M, Söll, D (2005) The complete set of tRNA species in Nanoarchaeum equitans. *FEBS Lett.* 579:2945–2947.

[21] Fujishima, K, Sugahara, J, Kikuta, K, Hirano, R, Sato, A, Tomita, M, Kanai, A (2009) Tri-split tRNA is a transfer RNA made from 3 transcripts that provides insight into the evolution of fragemented tRNAs in archaea. *Proc. Natl. Acad. Sci. USA* 106:2683–2687.

[22] Randau, L, Stanley, B, Kohlway, A, Mechta, S, Xiong, Y, Söll, D (2009) A cytidine deaminase edits C to U in transfer RNAs in archaea. *Science* 324:657–659.

[23] Gu, W, Jackman, J, Lohan, A, Gray, M, Phizicky, E (2003) tRNAHis maturation: An essential yeast protein catalyzes addition of a guanine nucleotide to the 5' end of tRNAHis. *Genes Dev.* 17:2889–2901.

[24] Jackman, J, Phizicky, E (2008) Identification of critical residues for G-1 addition and substrate recognition by tRNA(His) guanylyltransferase. *Biochemistry* 47:4817–4825.

[25] Heinemann, I, O'Donoghue, P, Madinger, C, Benner, J, Randau, L, Noren, C, Söll, D (2010) The appearance of pyrrolysine in tRNAHis guanylyltransferase by neutral evolution. *Proc. Natl. Acad. Sci. USA*.

[26] Schimmel, P, Yang, XL (2004) Two classes give lessons about CCA. *Nat. Struct. Mol. Biol.* 11:807–808 News.

[27] Xiong, Y, Steitz, T (2006) A story with a good ending: tRNA 3'-end maturation by CCA-adding enzymes. *Curr. Opin. Struct. Biol.* 16:12–17.

[28] Phizicky, EM, Alfonzo, JD (2010) Do all modifications benefit all tRNAs? *FEBS Lett.* 584:265–271.

[29] Martin, NC, Hopper, AK (1994) How single genes provide tRNA processing enzymes to mitochondria, nuclei and the cytosol. *Biochimie* 76:1161–1167.

[30] Söll, D, RajBhandary, UL (2006) The genetic code - thawing the 'frozen accident'. *J. Biosci.* 31:459–463.

[31] Muramatsu, T, Nishikawa, K, Nemoto, F, Kuchino, Y, Nishimura, S, Miyazawa, T, Yokoyama, S (1988) Codon and amino-acid specificities of a transfer RNA are both converted by a single post-transcriptional modification. *Nature* 336:179–181.

[32] Muramatsu, T, Yokoyama, S, Horie, N, Matsuda, A, Ueda, T, Yamaizumi, Z, Kuchino, Y, Nishimura, S, Miyazawa, T (1988) A novel lysine-substituted nucleoside in the first position of the anticodon of minor isoleucine tRNA from Escherichia coli. *J. Biol. Chem.* 263:9261–9267.

[33] Ikeuchi, Y, Soma, A, Ote, T, Kato, J, Sekine, Y, Suzuki, T (2005) molecular mechanism of lysidine synthesis that determines tRNA identity and codon recognition. *Mol. Cell* 19:235–246.

[34] Sampson, JR, Uhlenbeck, OC (1988) Biochemical and physical characterization of an unmodified yeast phenylalanine transfer RNA transcribed in vitro. *Proc. Natl. Acad. Sci. U.S.A.* 85:1033–1037.

[35] Hall, KB, Sampson, JR, Uhlenbeck, OC, Redfield, AG (1989) Structure of an unmodified tRNA molecule. *Biochemistry* 28:5794–5801.

[36] Perret, V, Garcia, A, Puglisi, J, Grosjean, H, Ebel, JP, Florentz, C, Giege, R (1990) Conformation in solution of yeast tRNA(Asp) transcripts deprived of modified nucleotides. *Biochimie* 72:735–743.

[37] Derrick, WB, Horowitz, J (1993) Probing structural differences between native and in vitro transcribed Escherichia coli valine transfer RNA: evidence for stable base modification-dependent conformers. *Nucleic Acids Res.* 21:4948–4953.

[38] Yue, D, Kintanar, A, Horowitz, J (1994) Nucleoside modifications stabilize Mg2+ binding in Escherichia coli tRNA(Val): an imino proton NMR investigation. *Biochemistry* 33:8905–8911.

[39] Maglott, EJ, Deo, SS, Przykorska, A, Glick, GD (1998) Conformational transitions of an unmodified tRNA: implications for RNA folding. *Biochemistry* 37:16349–16359.

[40] Serebrov, V, Vassilenko, K, Kholod, N, Gross, HJ, Kisselev, L (1998) Mg2+ binding and structural stability of mature and in vitro synthesized unmodified Escherichia coli tRNAPhe. *Nucleic Acids Res.* 26:2723–2728.

[41] Vermeulen, A, McCallum, SA, Pardi, A (2005) Comparison of the global structure and dynamics of native and unmodified tRNAval. *Biochemistry* 44:6024–6033.

[42] Nobles, KN, Yarian, CS, Liu, G, Guenther, RH, Agris, PF (2002) Highly conserved modified nucleosides influence Mg2+-dependent tRNA folding. *Nucleic Acids Res.* 30:4751–4760.

[43] Kawai, G, Yamamoto, Y, Kamimura, T, Masegi, T, Sekine, M, Hata, T, Iimori, T, Watanabe, T, Miyazawa, T, Yokoyama, S (1992) Conformational rigidity of specific pyrimidine residues in tRNA arises from posttranscriptional modifications that enhance steric interaction between the base and the 2'-hydroxyl group. *Biochemistry* 31:1040–1046.

[44] Drake, AF, Mason, SF, Trim, AR (1974) Optical studies of the base-stacking properties of 2'-O-methylated dinucleoside monophosphates. *J. Mol. Biol.* 86:727–739.

[45] Zmudzka, B, Bollum, FJ, Shugar, D (1969) Polydeoxyribouridylic acid and its complexes with polyribo- and deoxyriboadenylic acids. *J. Mol. Biol.* 46:169–183.

[46] Durant, P, Davis, D (1999) Stabilization of the anticodon stem-loop of tRNALys, 3 by an A+-C base-pair and by pseudouridine. *J. Mol. Biol.* 285:115–131.

[47] Yarian, C, Basti, M, Cain, R, Ansari, G, Guenther, R, Sochacka, E, Czerwinska, G, Malkiewicz, A, Agris, P (1999) Structural and functional roles of the N1- and N3-protons of psi at tRNA's position 39. *Nucl. Acids Res.* 27:3543–3549.

[48] Newby, M, Greenbaum, N (2001) A conserved pseudouridine modification in eukaryotic U2 snRNA induces a change in branch-site architecture. *RNA* 7:833–845.

[49] Newby, M, Greenbaum, N (2002) Sculpting of the spliceosomal branch site recognition motif by a conserved pseudouridine. *Nat. Struct. Biol.* 9:958–965.

[50] Dalluge, JJ, Hashizume, T, Sopchik, AE, McCloskey, JA, Davis, DR (1996) Conformational flexibility in RNA: the role of dihydrouridine. *Nucl. Acids Res.* 24:1073–1079.

[51] Schimmel, P, Redfield, A (1980) Transfer RNA in solution: selected topics. *Annu. Rev. Biophys. Bioeng.* 9:181–221.

[52] Jack, A, Ladner, J, Rhodes, D, Brown, R, Klug, A (1977) A crystallographic study of metal-binding to yeast phenylalanine transfer RNA. *J. Mol. Biol.* 111:315–328.

[53] Quigley, GJ, Teeter, MM, Rich, A (1978) Structural analysis of spermine and magnesium ion binding to yeast phenylalanine transfer RNA. *Proc. Natl. Acad. Sci. USA* 75:64–68.

[54] Shi, H, Moore, P (2000) The crystal structure of yeast phenylalanine tRNA at 1.93 Angstrom resolution: a classic structure revisited. *RNA* 6:1091–1105.

[55] Misra, V, Draper, D (2000) Mg(2+) binding to tRNA revisited: the nonlinear Poisson-Boltzmann model. *J. Mol. Biol.* 299:813–825.

[56] Draper, D (2004) A guide to ions and RNA structure. *RNA* 10:335–343.

[57] Jovine, L, Djordjevic, S, Rhodes, D (2000) The crystal structure of yeast phenylalanine tRNA at 2.0 A resolution: cleavage by Mg(2+) in 15-year old crystals. *J. Mol. Biol.* 301:401–414.

[58] Mikkelsen, NE, Johansson, K, Virtanen, A, Kirsebom, LA (2001) Aminoglycoside binding displaces a divalent metal ion in a tRNA-neomycin B complex. *Nat. Struct. Biol.* 8:510–514.

[59] Nakama, T, Nureki, O, Yokoyama, S (2001) Structural basis for the recognition of isoleucyl-adenylate and an antibiotic, mupirocin, by isoleucyl-tRNA synthetase. *J. Biol. Chem.* 276:47387–47393.

[60] Kobayashi, T, Nureki, O, Ishitani, R, Yaremchuk, A, Tukalo, M, Cusack, S, Sakamoto, K, Yokoyama, S (2003) Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion. *Nat. Struct. Biol.* 10:425–432.

[61] Chimnaronk, S, Forouhar, F, Sakai, J, Yao, M, Tron, CM, Atta, M, Fontecave, M, Hunt, JF, Tanaka, I (2009) Snapshots of dynamics in synthesizing N(6)-isopentenyladenosine at the tRNA anticodon. *Biochemistry* 48:5057–5065.

[62] Goto-Ito, S, Ito, T, Kuratani, M, Bessho, Y, Yokoyama, S (2009) Tertiary structure checkpoint at anticodon loop modification in tRNA functional maturation. *Nat. Struct. Mol. Biol.* 16:1109–1115.

[63] Eargle, J, Black, A, Sethi, A, Trabuco, L, Luthey-Schulten, ZA (2008) Dynamics of Recognition between tRNA and Elongation Factor Tu. *J. Mol. Biol.* 377:1382–1405.

[64] Woese, CR, Olsen, GJ, Ibba, M, Söll, D (2000) Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiol. Mol. Biol. Rev.* 64:202–236.

[65] Giege, R, Sissler, M, Florentz, C (1998) Universal rules and idiosyncratic features in tRNA identity. *Nucl. Acids Res.* 26:5017–5035.

[66] Ardell, D (2010) Computational analysis of tRNA identity. *FEBS Lett.* 584:325–333.

[67] Vasil'eva, I, Moor, N (2007) Interaction of aminoacyl-tRNA synthetases with tRNA: general principles and distinguishing characteristics of the high-molecular-weight substrate recognition. *Biochemistry Mosc.* 72:247–263.

[68] Ibba, M, Söll, D (2000) Aminoacyl-tRNA synthesis. *Ann. Rev. Biochem* 69:617–650.

[69] Arnez, J, Moras, D (1997) Structural and functional considerations of the aminoacylation reaction. *Trends Biochem. Sci.* 22:211–216.

[70] O'Donoghue, P, Luthey-Schulten, Z (2003) On the evolution of structure in the aminocyl-tRNA synthetases. *Microbiol. Mol. Bio. Rev.* 67:550–573.

[71] Perona, J, Hou, Y (2007) Indirect readout of tRNA for aminoacylation. *Biochemistry* 46:10419–10432.

[72] Uter, NT, Perona, JJ (2004) Long-range intramolecular signaling in a tRNA synthetase complex revealed by pre-steady-state kinetics. *Proc. Natl. Acad. Sci. U.S.A.* 101:14396–14401.

[73] Fersht, AR, Gangloff, J, Dirheimer, G (1978) Reaction pathway and rate-determining step in the aminoacylation of tRNAArg catalyzed by the arginyl-tRNA synthetase from yeast. *Biochemistry* 17:3740–3746.

[74] Eldred, EW, Schimmel, PR (1972) Investigation of the transfer of amino acid from a transfer ribonucleic acid synthetase-aminoacyl adenylate complex to transfer ribonucleic acid. *Biochemistry* 11:17–23.

[75] Yarus, M, Berg, P (1969) Recognition of tRNA by isoleucyl-tRNA synthetase. Effect of substrates on the dynamics of tRNA-enzyme interaction. *J. Mol. Biol.* 42:171–189.

[76] Zhang, C, Perona, J, Ryu, K, Francklyn, C, Hou, Y (2006) Distinct kinetic mechanisms of the two classes of Aminoacyl-tRNA synthetases. *J. Mol. Biol.* 361:300–311.

[77] Reed, VS, Wastney, ME, Yang, DC (1994) Mechanisms of the transfer of aminoacyl-tRNA from aminoacyl-tRNA synthetase to the elongation factor 1 alpha. *J. Biol. Chem.* 269:32932–32936.

[78] Hausmann, C, Praetorius-Ibba, M, Ibba, M (2007) An aminoacyl-tRNA synthetase:elongation factor complex for substrate channeling in archaeal translation. *Nucl. Acids Res.* 35:6094–6102.

[79] Guth, E, Connolly, SH, Bovee, M, Francklyn, CS (2005) A substrate-assisted concerted mechanism for aminoacylation by a class II aminoacyl-tRNA synthetase. *Biochemistry* 44:3785–3794.

[80] Dibbelt, L, Pachmann, U, Zachau, HG (1980) Serine activation is the rate limiting step of tRNASer aminoacylation by yeast seryl tRNA synthetase. *Nucleic Acids Res.* 8:4021–4039.

[81] Lin, SX, Baltzinger, M, Remy, P (1983) Fast kinetic study of yeast phenylalanyl-tRNA synthetase: an efficient discrimination between tyrosine and phenylalanine at the level of the aminoacyladenylate-enzyme complex. *Biochemistry* 22:681–689.

[82] Dibbelt, L, Zachau, HG (1981) On the rate limiting step of yeast tRNAPhe aminoacylation. *FEBS Lett.* 129:173–176.

[83] Francin, M, Kaminska, M, Kerjan, P, Merande, M (2002) The N-terminal domain of mammalian lysyl-tRNA synthetase is a functional tRNA–binding domain. *J. Biol. Chem.* 277:1762–1769.

[84] Kaminska, M, Shalak, V, Mirande, M (2001) The appended C–domain of human methionyl-tRNA synthetase has a tRNA–sequestering function. *Biochemistry* 40:14309–14316.

[85] Guo, M, Schimmel, P, Yang, XL (2010) Functional expansion of human tRNA synthetases achieved by structural inventions. *FEBS Lett.* 584:434–442.

[86] Hausmann, CD, Ibba, M (2008) Aminoacyl-tRNA synthetase complexes: molecular multitasking revealed. *FEMS Microbiol. Rev.* 32:705–721.

[87] Karkhanis, V, Mascarenhas, A, Martinis, S (2007) Amino acid toxicities of *Escherichia coli* that are prevented by leucyl-tRNA synthetase amino acid editing. *J. Bacteriol.* 189:8765–8768.

[88] Lee, J, Beebe, K, Nangle, L, Jang, J, Longo-Guess, C, Cook, S, Davisson, M, Sundberg, J, Schimmel, P, Ackerman, S (2006) Editing-defective tRNA synthetase causes protein misfolding and neurodegeneration. *Nature* 443:50–55.

[89] (2008) *Fidelity mechanisms of the aminoacyl-tRNA synethetases in: Protein Engineering* (Springer-Verlag), pp 153–200.

[90] Francklyn, C, Minajigi, A (2010) tRNA as an active chemical scaffold for diverse chemical transformations. *FEBS Lett.* 584:366–375.

[91] Sauerwald, A, Zhu, W, Major, TA, Roy, H, Palioura, S, Jahn, D, Whitman, WB, Yates, JRr, Ibba, M, Söll, D (2005) RNA-dependent cysteine biosynthesis in archaea. *Science* 307:1969–1972.

[92] Louie, A, Ribeiro, N, Reid, B, Jurnak, F (1984) Relative affinities of all Escherichia coli aminoacyl-tRNAs for elongation factor Tu-GTP. *J. Biol. Chem.* 259:5010–5016.

[93] Ott, G, Schiesswohl, M, Kiesewetter, S, Forster, C, Arnold, L, Erdmann, V, Sprinzl, M (1990) Ternary complexes of Escherichia coli aminoacyl-tRNAs with the elongation factor Tu and GTP: thermodynamic and structural studies. *Biochim. Biophys. Acta* 1050:222–225.

[94] Abrahamson, J, Laue, T, Miller, D, Johnson, A (1985) Direct determination of the association constant between elongation factor Tu X GTP and aminoacyl-tRNA using fluorescence. *Biochemistry* 24:692–700.

[95] Janiak, F, Dell, V, Abrahamson, J, Watson, B, Miller, D, Johnson, A (1990) Fluorescence characterization of the interaction of various transfer RNA species with elongation factor Tu·GTP: evidence for a new functional role for elongation factor Tu in protein biosynthesis. *Biochemistry* pp 4268–4277.

[96] LaRiviere, FJ, Wolfson, AD, Uhlenbeck, OC (2001) Uniform binding of aminoacyl-tRNAs to elongation factor Tu by thermodynamic compensation. *Science* 294:165–168.

[97] Voorhees, R, Weixlbaumer, A, Loakes, D, Kelley, A, Ramakrishnan, V (2009) Insights into substrate stabilization from snapshots of the peptidyl transferase center of the intacts 70S ribosome. *Nat. Struct. Mol. Biol.* 16:528–533.

[98] Ling, J, So, B, Yadavalli, S, Roy, H, Shoji, S, Fredrick, K, Musier-Forsyth, K, Ibba, M (2009) Resampling and editing mischarged tRNA prior to translation elongation. *Mol. Cell* 33:654–660.

[99] Nissen, P, Kjeldgaard, M, Thirup, S, Polekhina, G, Reshetnikova, L, Clark, BFC, Nyborg, J (1995) Crystal structure of the ternary complex of Phe-tRNA$^{\mathrm{Phe}}$, EF-Tu, and a GTP analog. *Science* 270:1464–1472.

[100] Nissan, TA, Oliphant, B, Perona, JJ (1999) An engineered class I transfer RNA with a class II tertiary fold. *RNA* 5:434–445.

[101] Asahara, H, Uhlenbeck, OC (2002) The tRNA specificity of Thermus thermophilus EF-Tu. *Proc. Natl. Acad. Sci. USA* 99:3499–3504.

[102] Dale, T, Sanderson, LE, Uhlenbeck, OC (2004) The affinity of elongation factor Tu for an aminoacyl-tRNA is modulated by the esterified amino acid. *Biochemistry* 43:6159–6166.

[103] Korostelev, A, Trakhanov, S, Laurberg, M, Noller, H (2006) Crystal structure of a 70S ribosome-tRNA complex reveals functional interactions and rearrangements. *Cell* 126:1065–1077.

[104] Selmer, M, Dunham, C, Murthy, F, Weixlbaumer, A, Petry, S, Kelley, A, Weir, J, Ramakrishnan, V (2006) Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* 313:1935–1942.

[105] Khade, P, Joseph, S (2010) Functional interactions by transfer RNAs in the ribosome. *FEBS Lett.* 584:420–426.

[106] Agirrezabala, X, Frank, J (2009) Elongation in translation as a dynamic interaction among the ribosome, tRNA, and elongation factors EF-G and EF-Tu. *Quart. Rev. Biophys.* 42:159–200.

[107] Ban, N, Nissen, P, Hansen, J, Moore, PB, Steitz, TA (2000) The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science* 289:905–920.

[108] Nissen, P, Hansen, J, Ban, N, Moore, P, Steitz, T (2000) The structural basis of ribosome activity in peptide bond synthesis. *Science* 289:920–930.

[109] Harms, J, Schluenzen, F, Zarivach, R, Bashan, A, Gat, S, Agmon, I, Bartels, H, Franceschi, F, Yonath, A (2001) High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. *Cell* 107:679–688.

[110] Moazed, D, Noller, H (1989) Intermediate states in the movement of transfer RNA in the ribosome. *Nature* 342:142–148.

[111] Valle, M, Zavialov, A, Li, W, Stagg, S, Sengupta, J, Nielsen, R, Nissen, P, Harvey, S, Ehrenberg, M, Frank, J (2003) Incorporation of aminoacyl-tRNA into the ribosome as seen by cryo-electron microscopy. *Nat. Struct. Biol.* 10:899–906.

[112] Pape, T, Wintermeye, W, Rodnina, M (1999) Induced fit in initial selection and proofreading of aminoacyl-tRNA on the ribosome. *EMBO J.* 18:3800–3807.

[113] Rodnina, M, Savelsbergh, A, Katunin, V, Wintermeyer, W (1997) Hydrolysis of GTP by elongation factor G drives tRNA movement on the ribosome. *Nature* 385:37–41.

[114] Ogle, JM, Brodersen, DE, Clemons, WM, Tarry, MJ, Carter, AP, Ramakrishnan, V (2001) Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* 292:897–902.

[115] Fahlman, RP, Dale, T, Uhlenbeck, OC (2004) Uniform binding of aminoacylated transfer RNAs to the ribosomal A and P sites. *Mol. Cell* 16:799–805.

[116] Ledoux, S, Uhlenbeck, OC (2008) Different aa-tRNAs are selected uniformly on the ribosome. *Mol. Cell* 31:114–123.

[117] Yarus, M, Cline, SW, Wier, P, Breeden, L, Thompson, RC (1986) Actions of the anticodon arm in translation on the phenotypes of RNA mutants. *J. Mol. Biol.* 192:235–255.

[118] Yarus, M, Cline, S, Raftery, L, Wier, P, Bradley, D (1986) The translational efficiency of tRNA is a property of the anticodon arm. *J. Biol. Chem.* 261:10496–10505.

[119] Hou, YM, Schimmel, P (1988) A simple structural feature is a major determinant of the identity of a transfer RNA. *Nature* 333:140–145.

[120] Kleina, LG, Masson, JM, Normanly, J, Abelson, J, Miller, JH (1990) Construction of Escherichia coli amber suppressor tRNA genes. II. Synthesis of additional tRNA genes and improvement of suppressor efficiency. *J. Mol. Biol.* 213:705–717.

[121] Schultz, DW, Yarus, M (1994) tRNA structure and ribosomal function. II. Interaction between anticodon helix and other tRNA mutations. *J. Mol. Biol.* 235:1395–1405.

[122] Schultz, DW, Yarus, M (1994) tRNA structure and ribosomal function. I. tRNA nucleotide 27-43 mutations enhance first position wobble. *J. Mol. Biol.* 235:1381–1394.

[123] McClain, WH, Schneider, J, Bhattacharya, S, Gabriel, K (1998) The importance of tRNA backbone-mediated interactions with synthetase for aminoacylation. *Proc. Natl. Acad. Sci. U.S.A.* 95:460–465.

[124] Takai, K, Takaku, H, Yokoyama, S (1996) Codon-reading specificity of an unmodified form of Escherichia coli tRNA1Ser in cell-free protein synthesis. *Nucleic Acids Res.* 24:2894–2899.

[125] Kruger, MK, Pedersen, S, Hagervall, TG, S?rensen, MA (1998) The modification of the wobble base of tRNAGlu modulates the translation rate of glutamic acid codons in vivo. *J. Mol. Biol.* 284:621–631.

[126] Konevega, AL, Soboleva, NG, Makhno, VI, Semenkov, YP, Wintermeyer, W, Rodnina, MV, Katunin, VI (2004) Purine bases at position 37 of tRNA stabilize codon-anticodon interaction in the ribosomal A site by stacking and Mg2+-dependent interactions. *RNA* 10:90–101.

[127] Murphy, FV, Ramakrishnan, V, Malkiewicz, A, Agris, PF (2004) The role of modifications in codon discrimination by tRNA(Lys)UUU. *Nat. Struct. Mol. Biol.* 11:1186–1191.

[128] Yamada, Y, Matsugi, J, Ishikura, H, Murao, K (2005) Bacillus subtilis tRNA(Pro) with the anticodon mo5UGG can recognize the codon CCC. *Biochim. Biophys. Acta* 1728:143–149.

[129] Cochella, L, Green, R (2005) An active role for tRNA in decoding beyond codon:anticodon pairing. *Science* 308:1178–1180.

[130] Murakami, H, Ohta, A, Suga, H (2009) Bases in the anticodon loop of tRNA(Ala)(GGC) prevent misreading. *Nat. Struct. Mol. Biol.* 16:353–358.

[131] Sethi, A, Eargle, J, Black, A, Luthey-Schulten, ZA (2009) Dynamical networks in tRNA:protein complexes. *Proc. Natl. Acad. Sci. USA* 106:6620–6625.

[132] Black Pyrkosz, A, Eargle, J, Sethi, A, Luthey-Schulten, Z (2010) Exit strategies for charged tRNA from GluRS. *J. Mol. Biol.* 397:1350–1371.

[133] Alexander, RW, Eargle, J, Luthey-Schulten, Z (2010) Experimental and computational determination of tRNA dynamics. *FEBS Lett.* 584:376–386.

[134] Krab, IM, Parmeggiani, A (1998) EF-Tu, a GTPase odyssey. *Biochim. Biophys. Acta* 1443:1–22.

[135] Lowe, TM, Eddy, SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucl. Acids Res.* 25:955–964.

[136] Markowitz, VM, Chen, IM, Palaniappan, K, Chu, K, Szeto, E, Grechkin, Y, Ratner, A, Anderson, I, Lykidis, A, Mavromatis, K, Ivanova, NN, Kyrpides, NC (2010) The integrated microbial genomes system: an expanding comparative analysis resource. *Nucleic Acids Res.* 38:D382–390.

[137] Li, W, Jaroszewski, L, Godzik, A (2002) Tolerating some redundancy significantly speeds up clustering of large protein databases. *Bioinformatics* 18:77–82.

[138] Park, J, Holm, L, Heger, A, Chothia, C (2000) RSDB: representative protein sequence databases have high information content. *Bioinformatics* 16:458–464.

[139] Sethi, A, O'Donoghue, P, Luthey-Schulten, Z (2005) Evolutionary profiles from the qr factorization of multiple sequence alignments. *Proc. Natl. Acad. Sci. USA* 102:4045–4050.

[140] Stombaugh, J, Zirbel, CL, Westhof, E, Leontis, NB (2009) Frequency and isostericity of RNA base pairs. *Nucleic Acids Res.* 37:2294–2312.

[141] Zirbel, CL, Sponer, JE, Sponer, J, Stombaugh, J, Leontis, NB (2009) Classification and energetics of the base-phosphate interactions in RNA. *Nucleic Acids Res.* 37:4898–4918.

[142] Shannon, C (1948) A mathematical theory of communication. *Bell System TEch. J.* 27:623–656.

[143] Schneider, TD, Stormo, GD, Gold, L, Ehrenfeucht, A (1986) Information content of binding sites on nucleotide sequences. *J. Mol. Biol.* 188:415–431.

[144] Cavener, DR (1987) Comparison of the consensus sequence flanking translational start sites in Drosophila and vertebrates. *Nucleic Acids Res* 15:1353–1361.

[145] Markowitz, VM, Korzeniewski, F, Palaniappan, K, Szeto, E, Werner, G, Padki, A, Zhao, X, Dubchak, I, Hugenholtz, P, Anderson, I, Lykidis, A, Mavromatis, K, Ivanova, N, Kyrpides, NC (2006) The integrated microbial genomes (IMG) system. *Nucl. Acids Res.* 34:344–348.

[146] Juhling, F, Morl, M, Hartmann, RK, Sprinzl, M, Stadler, PF, Putz, J (2009) tRNAdb 2009: compilation of tRNA sequences and tRNA genes. *Nucleic Acids Res.* 37:D159–162.

[147] Abe, T, Ikemura, T, Ohara, Y, Uehara, H, Kinouchi, M, Kanaya, S, Yamada, Y, Muto, A, Inokuchi, H (2009) tRNADB-CE: tRNA gene database curated manually by experts. *Nucleic Acids Res.* 37:D163–168.

[148] Chan, PP, Lowe, TM (2009) GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res.* 37:D93–97.

[149] Thompson, JD, Higgins, DG, Gibson, TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* 22:4673–4680.

[150] Roberts, E, Eargle, J, Wright, D, Luthey-Schulten, Z (2006) MultiSeq: unifying sequence and structure data for evolutionary analysis. *BMC Bioinformatics* 7:382.

[151] Humphrey, W, Dalke, A, Schulten, K (1996) VMD: visual molecular dynamics. *J. Mol. Graphics* 14:33–38.

[152] Francklyn, C, Musier-Forsyth, K, Schimmel, P (1992) Small RNA helices as substrates for aminoacylation and their relationship to charging of transfer RNAs. *Eur. J. Biochem.* 206:315–321.

[153] Park, SJ, Schimmel, P (1988) Evidence for interaction of an aminoacyl transfer RNA synthetase with a region important for the identity of its cognate transfer RNA. *J. Biol. Chem.* 263:16527–16530.

[154] Nissan, TA, Perona, JJ (2000) Alternative designs for construction of the class II transfer RNA tertiary core. *RNA* 6:1585–1596.

[155] Pallanck, L, Li, S, Schulman, LH (1992) The anticodon and discriminator base are major determinants of cysteine tRNA identity in vivo. *J. Biol. Chem.* 267:7221–7223.

[156] Ming, X, Smith, K, Suga, H, Hou, YM (2002) Recognition of tRNA backbone for aminoacylation with cysteine: evolution from Escherichia coli to human. *J. Mol. Biol.* 318:1207–1220.

[157] Christian, T, Lipman, RS, Evilia, C, Hou, YM (2000) Alternative design of a tRNA core for aminoacylation. *J. Mol. Biol.* 303:503–514.

[158] Wang, Y, Rader, AJ, Bahar, I, Jernigan, RL (2004) Global ribosome motions revealed with elastic network model. *J. Struct. Biol.* 147:302–314.

[159] McClain, WH, Foss, K (1988) Changing the identity of a tRNA by introducing a G-U wobble pair near the 3' acceptor end. *Science* 240:793–796.

[160] Francklyn, C, Shi, JP, Schimmel, P (1992) Overlapping nucleotide determinants for specific aminoacylation of RNA microhelices. *Science* 255:1121–1125.

[161] Hou, YM, Schimmel, P (1989) Evidence that a major determinant for the identity of a transfer RNA is conserved in evolution. *Biochemistry* 28:6800–6804.

[162] Shi, JP, Francklyn, C, Hill, K, Schimmel, P (1990) A nucleotide that enhances the charging of RNA minihelix sequence variants with alanine. *Biochemistry* 29:3621–3626.

[163] Tamura, K, Asahara, H, Himeno, H, Hasegawa, T, Shimizu, M (1991) Identity elements of Escherichia coli tRNA(Ala). *J. Mol. Recognit.* 4:129–132.

[164] Gabriel, K, Schneider, J, McClain, WH (1996) Functional evidence for indirect recognition of G.U in tRNA(Ala) by alanyl-tRNA synthetase. *Science* 271:195–197.

[165] Imura, N, Weiss, GB, Chambers, RW (1969) Reconstitution of alanine acceptor activity from fragments of yeast tRNA-Ala II. *Nature* 222:1147–1148.

[166] Carneiro, VT, Dietrich, A, Marechal-Drouard, L, Cosset, A, Pelletier, G, Small, I (1994) Characterization of some major identity elements in plant alanine and phenylalanine transfer RNAs. *Plant Mol. Biol.* 26:1843–1853.

[167] Beuning, PJ, Yang, F, Schimmel, P, Musier-Forsyth, K (1997) Specific atomic groups and RNA helix geometry in acceptor stem recognition by a tRNA synthetase. *Proc. Natl. Acad. Sci. U.S.A.* 94:10150–10154.

[168] McClain, WH, Gabriel, K, Schneider, J (1996) Specific function of a G.U wobble pair from an adjacent helical site in tRNA(Ala) during recognition by alanyl-tRNA synthetase. *RNA* 2:105–109.

[169] Mueller, U, Schubel, H, Sprinzl, M, Heinemann, U (1999) Crystal structure of acceptor stem of tRNA(Ala) from Escherichia coli shows unique G.U wobble base pair at 1.16 A resolution. *RNA* 5:670–677.

[170] Choi, H, Otten, S, Schneider, J, McClain, WH (2002) Genetic perturbations of RNA reveal structure-based recognition in protein-RNA interaction. *J. Mol. Biol.* 324:573–576.

[171] Kallick, DA, Nagan, MC, Beuning, PJ, Kerimo, S, Tessmer, MR, Cramer, CJ, Musier-Forsyth, K (2002) Discrimination of C1:G72 microhelix(Ala) by AlaRS is based on specific atomic groups rather than conformational effects: An NMR and MD analysis. . *J. Phys. Chem. B* 106:8878–8884.

[172] Mallick, B, Chakrabarti, J, Sahoo, S, Ghosh, Z, Das, S (2005) Identity elements of archaeal tRNA. *DNA Res.* 12:235–246.

[173] McClain, WH, Foss, K (1988) Changing the acceptor identity of a transfer RNA by altering nucleotides in a "variable pocket". *Science* 241:1804–1807.

[174] Schulman, LH, Pelka, H (1989) The anticodon contains a major element of the identity of arginine transfer RNAs. *Science* 246:1595–1597.

[175] McClain, WH, Foss, K, Jenkins, RA, Schneider, J (1990) Nucleotides that determine Escherichia coli tRNA(Arg) and tRNA(Lys) acceptor identities revealed by analyses of mutant opal and amber suppressor tRNAs. *Proc. Natl. Acad. Sci. U.S.A.* 87:9260–9264.

[176] Tamura, K, Himeno, H, Asahara, H, Hasegawa, T, Shimizu, M (1992) In vitro study of E.coli tRNA(Arg) and tRNA(Lys) identity elements. *Nucleic Acids Res.* 20:2335–2339.

[177] Sissler, M, Giege, R, Florentz, C (1998) The RNA sequence context defines the mechanistic routes by which yeast arginyl-tRNA synthetase charges tRNA. *RNA* 4:647–657.

[178] Shimada, A, Nureki, O, Goto, M, Takahashi, S, Yokoyama, S (2001) Structural and mutational studies of the recognition of the arginine tRNA-specific major identity element, A20, by arginyl-tRNA synthetase. *Proc. Natl. Acad. Sci. U.S.A.* 98:13537–13542.

[179] Geslain, R, Martin, F, Camasses, A, Eriani, G (2003) A yeast knockout strain to discriminate between active and inactive tRNA molecules. *Nucleic Acids Res.* 31:4729–4737.

[180] Guigou, L, Mirande, M (2005) Determinants in tRNA for activation of arginyl-tRNA synthetase: evidence that tRNA flexibility is required for the induced-fit mechanism. *Biochemistry* 44:16540–16548.

[181] Hasegawa, T, Himeno, H, Ishikura, H, Shimizu, M (1989) Discriminator base of tRNA(Asp) is involved in amino acid acceptor activity. *Biochem. Biophys. Res. Commun.* 163:1534–1538.

[182] Nameki, N, Tamura, K, Himeno, H, Asahara, H, Hasegawa, T, Shimizu, M (1992) Escherichia coli tRNA(Asp) recognition mechanism differing from that of the yeast system. *Biochem. Biophys. Res. Commun.* 189:856–862.

[183] Putz, J, Puglisi, JD, Florentz, C, Giege, R (1991) Identity elements for specific aminoacylation of yeast tRNA(Asp) by cognate aspartyl-tRNA synthetase. *Science* 252:1696–1699.

[184] Frugier, M, Söll, D, Giege, R, Florentz, C (1994) Identity switches between tRNAs aminoacylated by class I glutaminyl- and class II aspartyl-tRNA synthetases. *Biochemistry* 33:9912–9921.

[185] Becker, HD, Giege, R, Kern, D (1996) Identity of prokaryotic and eukaryotic tRNA(Asp) for aminoacylation by aspartyl-tRNA synthetase from Thermus thermophilus. *Biochemistry* 35:7447–7458.

[186] Choi, H, Gabriel, K, Schneider, J, Otten, S, McClain, WH (2003) Recognition of acceptor-stem structure of tRNA(Asp) by Escherichia coli aspartyl-tRNA synthetase. *RNA* 9:386–393.

[187] Giege, R, Florentz, C, Kern, D, Gangloff, J, Eriani, G, Moras, D (1996) Aspartate identity of transfer RNAs. *Biochimie* 78:605–623.

[188] Fender, A, Geslain, R, Eriani, G, Giege, R, Sissler, M, Florentz, C (2004) A yeast arginine specific tRNA is a remnant aspartate acceptor. *Nucleic Acids Res.* 32:5076–5086.

[189] Shimizu, M, Asahara, H, Tamura, K, Hasegawa, T, Himeno, H (1992) The role of anticodon bases and the discriminator nucleotide in the recognition of some E. coli tRNAs by their aminoacyl-tRNA synthetases. *J. Mol. Evol.* 35:436–443.

[190] Hou, YM, Westhof, E, Giege, R (1993) An unusual RNA tertiary interaction has a role for the specific aminoacylation of a transfer RNA. *Proc. Natl. Acad. Sci. U.S.A.* 90:6776–6780.

[191] McClain, WH (1993) Identity of Escherichia coli tRNA(Cys) determined by nucleotides in three regions of tRNA tertiary structure. *J. Biol. Chem.* 268:19398–19402.

[192] Komatsoulis, GA, Abelson, J (1993) Recognition of tRNA(Cys) by Escherichia coli cysteinyl-tRNA synthetase. *Biochemistry* 32:7435–7444.

[193] Hamann, C, Hou, Y (1997) An RNA structural determinant for tRNA recognition. *Biochemistry* 36:7967–7972.

[194] Hou, YM, Sterner, T, Bhalla, R (1995) Evidence for a conserved relationship between an acceptor stem and a tRNA for aminoacylation. *RNA* 1:707–713.

[195] Lipman, RS, Hou, YM (1998) Aminoacylation of tRNA in the evolution of an aminoacyl-tRNA synthetase. *Proc. Natl. Acad. Sci. USA* 95:13495–13500.

[196] Sherlin, LD, Bullock, TL, Newberry, KJ, Lipman, RS, Hou, YM, Beijer, B, Sproat, BS, Perona, JJ (2000) Influence of transfer RNA tertiary structure on aminoacylation efficiency by glutaminyl and cysteinyl-tRNA synthetases. *J. Mol. Biol.* 299:431–446.

[197] Shitivelband, S, Hou, YM (2005) Breaking the stereo barrier of amino acid attachment to tRNA by a single nucleotide. *J. Mol. Biol.* 348:513–521.

[198] Hou, YM, Motegi, H, Lipman, RS, Hamann, CS, Shiba, K (1999) Conservation of a tRNA core for aminoacylation. *Nucl. Acids Res.* 27:4743–4750.

[199] Evilia, C, Ming, X, DasSarma, S, Hou, YM (2003) Aminoacylation of an unusual tRNA(Cys) from an extreme halophile. *RNA* 9:794–801.

[200] Rogers, MJ, Söll, D (1988) Discrimination between glutaminyl-tRNA synthetase and seryl-tRNA synthetase involves nucleotides in the acceptor helix of tRNA. *Proc. Natl. Acad. Sci. U.S.A.* 85:6627–6631.

[201] Jahn, M, Rogers, MJ, Söll, D (1991) Anticodon and acceptor stem nucleotides in tRNA(Gln) are major recognition elements for E. coli glutaminyl-tRNA synthetase. *Nature* 352:258–260.

[202] Hayase, Y, Jahn, M, Rogers, MJ, Sylvers, LA, Koizumi, M, Inoue, H, Ohtsuka, E, Söll, D (1992) Recognition of bases in Escherichia coli tRNA(Gln) by glutaminyl-tRNA synthetase: a complete identity set. *EMBO J.* 11:4159–4165.

[203] Ibba, M, Hong, KW, Sherman, JM, Sever, S, Söll, D (1996) Interactions between tRNA identity nucleotides and their recognition sites in glutaminyl-tRNA synthetase determine the cognate amino acid affinity of the enzyme. *Proc. Natl. Acad. Sci. U.S.A.* 93:6953–6958.

[204] Fukunaga, R, Yokoyama, S (2007) Structural insights into the second step of RNA-dependent cysteine biosynthesis in archaea: crystal structure of Sep-tRNA:Cys-tRNA synthase from Archaeoglobus fulgidus. *J. Mol. Biol.* 370:128–141.

[205] Normanly, J, Kleina, LG, Masson, JM, Abelson, J, Miller, JH (1990) Construction of Escherichia coli amber suppressor tRNA genes. III. Determination of tRNA specificity. *J. Mol. Biol.* 213:719–726.

[206] Sylvers, LA, Rogers, KC, Shimizu, M, Ohtsuka, E, Söll, D (1993) A 2-thiouridine derivative in tRNAGlu is a positive determinant for aminoacylation by Escherichia coli glutamyl-tRNA synthetase. *Biochemistry* 32:3836–3841.

[207] Gregory, ST, Dahlberg, AE (1995) Effects of mutations at position 36 of tRNA(Glu) on missense and nonsense suppression in Escherichia coli. *FEBS Lett.* 361:25–28.

[208] Sekine, S, Nureki, O, Sakamoto, K, Niimi, T, Tateno, M, Gō, M, Kohno, T, Brisson, A, Lapointe, J, Yokoyama, S (1996) Major Identity Determinants in the "Augmented D Helix" of $tRNA^{Glu}$ from Escherichia coli. *J. Mol. Biol.* 256:685–700.

[209] Kruger, MK, S?rensen, MA (1998) Aminoacylation of hypomodified tRNAGlu in vivo. *J. Mol. Biol.* 284:609–620.

[210] McClain, WH, Foss, K, Jenkins, RA, Schneider, J (1991) Rapid determination of nucleotides that define tRNA(Gly) acceptor identity. *Proc. Natl. Acad. Sci. U.S.A.* 88:6147–6151.

[211] Francklyn, C, Shi, JP, Schimmel, P (1992) Overlapping nucleotide determinants for specific aminoacylation of RNA microhelices. *Science* 255:1121–1125.

[212] Nameki, N, Tamura, K, Asahara, H, Hasegawa, T (1997) Recognition of tRNA(Gly) by three widely diverged glycyl-tRNA synthetases. *J. Mol. Biol.* 268:640–647.

[213] Mazuric, MH (1997) *PhD Thesis* (Universite Louis Pasteur, Strasbourg).

[214] Hipps, D, Shiba, K, Henderson, B, Schimmel, P (1995) Operational RNA code for amino acids: species-specific aminoacylation of minihelices switched by a single nucleotide. *Proc. Natl. Acad. Sci. U.S.A.* 92:5550–5552.

[215] Mazauric, MH, Roy, H, Kern, D (1999) tRNA glycylation system from Thermus thermophilus. tRNAGly identity and functional interrelation with the glycylation systems from other phylae. *Biochemistry* 38:13094–13105.

[216] Himeno, H, Hasegawa, T, Ueda, T, Watanabe, K, Miura, K, Shimizu, M (1989) Role of the extra G-C pair at the end of the acceptor stem of tRNA(His) in aminoacylation. *Nucleic Acids Res.* 17:7855–7863.

[217] Yan, W, Augustine, J, Francklyn, C (1996) A tRNA identity switch mediated by the binding interaction between a tRNA anticodon and the accessory domain of a class II aminoacyl-tRNA synthetase. *Biochemistry* 35:6559–6568.

[218] Rudinger, J, Florentz, C, Giege, R (1994) Histidylation by yeast HisRS of tRNA or tRNA-like structure relies on residues -1 and 73 but is dependent on the RNA context. *Nucleic Acids Res.* 22:5031–5037.

[219] Nameki, N, Asahara, H, Shimizu, M, Okada, N, Himeno, H (1995) Identity elements of Saccharomyces cerevisiae tRNA(His). *Nucleic Acids Res.* 23:389–394.

[220] Hawko, SA, Francklyn, CS (2001) Covariation of a specificity-determining structural motif in an aminoacyl-tRNA synthetase and a tRNA identity element. *Biochemistry* 40:1930–1936.

[221] Connolly, SA, Rosen, AE, Musier-Forsyth, K, Francklyn, CS (2004) G-1:C73 recognition by an arginine cluster in the active site of Escherichia coli histidyl-tRNA synthetase. *Biochemistry* 43:962–969.

[222] Rosen, AE, Musier-Forsyth, K (2004) Recognition of G-1:C73 atomic groups by Escherichia coli histidyl-tRNA synthetase. *J. Am. Chem. Soc.* 126:64–65.

[223] Rosen, AE, Brooks, BS, Guth, E, Francklyn, CS, Musier-Forsyth, K (2006) Evolutionary conservation of a functionally important backbone phosphate group critical for aminoacylation of histidine tRNAs. *RNA* 12:1315–1322.

[224] Jackman, JE, Phizicky, EM (2006) tRNAHis guanylyltransferase adds G-1 to the 5' end of tRNAHis by recognition of the anticodon, one of several features unexpectedly shared with tRNA synthetases. *RNA* 12:1007–1014.

[225] Pallanck, L, Schulman, LH (1991) Anticodon-dependent aminoacylation of a noncognate tRNA with isoleucine, valine, and phenylalanine in vivo. *Proc. Natl. Acad. Sci. U.S.A.* 88:3872–3876.

[226] (1993) *The Translational Apparatus* (Plenu Press), pp 59–66.

[227] Nureki, O, Niimi, T, Muramatsu, T, Kanno, H, Kohno, T, Florentz, C, Giege, R, Yokoyama, S (1994) Molecular recognition of the identity-determinant set of isoleucine transfer RNA from Escherichia coli. *J. Mol. Biol.* 236:710–724.

[228] Normanly, J, Ollick, T, Abelson, J (1992) Eight base changes are sufficient to convert a leucine-inserting tRNA into a serine-inserting tRNA. *Proc. Natl. Acad. Sci. U.S.A.* 89:5680–5684.

[229] Asahara, H, Himeno, H, Tamura, K, Hasegawa, T, Watanabe, K, Shimizu, M (1993) Recognition nucleotides of Escherichia coli tRNA(Leu) and its elements facilitating discrimination from tRNASer and tRNA(Tyr). *J. Mol. Biol.* 231:219–229.

[230] Soma, A, Kumagai, R, Nishikawa, K, Himeno, H (1996) The anticodon loop is a major identity determinant of Saccharomyces cerevisiae tRNA(Leu). *J. Mol. Biol.* 263:707–714.

[231] Breitschopf, K, Gross, HJ (1994) The exchange of the discriminator base A73 for G is alone sufficient to convert human tRNA(Leu) into a serine-acceptor in vitro. *EMBO J.* 13:3166–3169.

[232] Breitschopf, K, Achsel, T, Busch, K, Gross, HJ (1995) Identity elements of human tRNA(Leu): structural requirements for converting human tRNA(Ser) into a leucine acceptor in vitro. *Nucleic Acids Res.* 23:3633–3637.

[233] Asahara, H, Nameki, N, Hasegawa, T (1998) In vitro selection of RNAs aminoacylated by Escherichia coli leucyl-tRNA synthetase. *J. Mol. Biol.* 283:605–618.

[234] Tocchini-Valentini, G, Saks, ME, Abelson, J (2000) tRNA leucine identity and recognition sets. *J. Mol. Biol.* 298:779–793.

[235] Du, X, Wang, E (2003) Tertiary structure base pairs between D- and TYC-loops of Escherichia coli tRNALeu play important roles in both aminoacylation and editing . *Nucl. Acids Res.* 31:2865–2872.

[236] Soma, A, Uchiyama, K, Sakamoto, T, Maeda, M, Himeno, H (1999) Unique recognition style of tRNA(Leu) by Haloferax volcanii leucyl-tRNA synthetase. *J. Mol. Biol.* 293:1029–1038.

[237] Yao, P, Zhu, B, Jaeger, S, Eriani, G, Wang, ED (2008) Recognition of tRNALeu by Aquifex aeolicus leucyl-tRNA synthetase during the aminoacylation and editing steps. *Nucleic Acids Res.* 36:2728–2738.

[238] Ibba, M, Losey, HC, Kawarabayasi, Y, Kikuchi, H, Bunjun, S, Söll, D (1999) Substrate recognition by class I lysyl-tRNA synthetases: a molecular basis for gene displacement. *Proc. Natl. Acad. Sci. USA* 96:418–423.

[239] Ambrogelly, A, Frugier, M, Ibba, M, Söll, D, Giege, R (2005) Transfer RNA recognition by class I lysyl-tRNA synthetase from the Lyme disease pathogen Borrelia burgdorferi. *FEBS Lett.* 579:2629–2634.

[240] Söll, D, Becker, HD, Plateau, P, Blanquet, S, Ibba, M (2000) Context-dependent anticodon recognition by class I lysyl-tRNA synthetases. *Proc. Natl. Acad. Sci. U.S.A.* 97:14224–14228.

[241] Ambrogelly, A, Korencic, D, Ibba, M (2002) Functional annotation of class I lysyl-tRNA synthetase phylogeny indicates a limited role for gene transfer. *J. Bacteriol.* 184:4594–4600.

[242] Shiba, K, Stello, T, Motegi, H, Noda, T, Musier-Forsyth, K, Schimmel, P (1997) Human lysyl-tRNA synthetase accepts nucleotide 73 variants and rescues Escherichia coli double-defective mutant. *J. Biol. Chem.* 272:22809–22816.

[243] Stello, T, Hong, M, Musier-Forsyth, K (1999) Efficient aminoacylation of tRNA(Lys,3) by human lysyl-tRNA synthetase is dependent on covalent continuity between the acceptor stem and the anticodon domain. *Nucleic Acids Res.* 27:4823–4829.

[244] Fukunaga, J, Ohno, S, Nishikawa, K, Yokogawa, T (2006) A base pair at the bottom of the anticodon stem is reciprocally preferred for discrimination of cognate tRNAs by Escherichia coli lysyl- and glutaminyl-tRNA synthetases. *Nucleic Acids Res.* 34:3181–3188.

[245] Bjork, GR, Huang, B, Persson, OP, Bystrom, AS (2007) A conserved modified wobble nucleoside (mcm5s2U) in lysyl-tRNA is required for viability in yeast. *RNA* 13:1245–1255.

[246] Francin, M, Mirande, M (2006) Identity elements for specific aminoacylation of a tRNA by mammalian lysyl-tRNA synthetase bearing a nonspecific tRNA-interacting factor. *Biochemistry* 45:10153–10160.

[247] Uemura, H, Imai, M, Ohtsuka, E, Ikehara, M, Söll, D (1982) E. coli initiator tRNA analogs with different nucleotides in the discriminator base position. *Nucleic Acids Res.* 10:6531–6539.

[248] Schulman, LH, Pelka, H (1988) Anticodon switching changes the identity of methionine and valine transfer RNAs. *Science* 242:765–768.

[249] Lee, C, RajBhandary, U (1991) Mutants of Escherichia coli initiator tRNA that suppress amber codons in Saccharomyces cerevisiae and are aminoacylated with tyrosine by yeast extracts. *Proc. Natl. Acad. Sci. U.S.A.* 88:11378–11382.

[250] Meinnel, T, Mechulam, Y, Lazennec, C, Blanquet, S, Fayat, G (1993) Critical role of the acceptor stem of tRNAs(Met) in their aminoacylation by Escherichia coli methionyl-tRNA synthetase. *J. Mol. Biol.* 229:26–36.

[251] Senger, B, Despons, L, Walter, P, Fasiolo, F (1992) The anticodon triplet is not sufficient to confer methionine acceptance to a transfer RNA. *Proc. Natl. Acad. Sci. U.S.A.* 89:10768–10771.

[252] Senger, B, Aphasizhev, R, Walter, P, Fasiolo, F (1995) The presence of a D-stem but not a T-stem is essential for triggering aminoacylation upon anticodon binding in yeast methionine tRNA. *J. Mol. Biol.* 249:45–58.

[253] Ramesh, V, RajBhandary, UL (2001) Importance of the anticodon sequence in the aminoacylation of tRNAs by methionyl-tRNA synthetase and by valyl-tRNA synthetase in an Archaebacterium. *J. Biol. Chem.* 276:3660–3665.

[254] Freyhult, E, Moulton, V, Ardell, DH (2006) Visualizing bacterial tRNA identity determinants and antideterminants using function logos and inverse function logos. *Nucleic Acids Res.* 34:905–916.

[255] McClain, WH, Foss, K (1988) Nucleotides that contribute to the identity of Escherichia coli tRNA(Phe). *J. Mol. Biol.* 202:697–709.

[256] Sampson, JR, DiRenzo, AB, Behlen, LS, Uhlenbeck, OC (1989) Nucleotides in yeast tRNAPhe required for the specific recognition by its cognate synthetase. *Science* 243:1363–1366.

[257] Sampson, JR, Behlen, LS, DiRenzo, AB, Uhlenbeck, OC (1992) Recognition of yeast tRNA(Phe) by its cognate yeast phenylalanyl-tRNA synthetase: an analysis of specificity. *Biochemistry* 31:4161–4167.

[258] Frugier, M, Florentz, C, Schimmel, P, Giege, R (1993) Triple aminoacylation specificity of a chimerized transfer RNA. *Biochemistry* 32:14053–14061.

[259] Moor, N, Nazarenko, I, Ankilova, V, Khodyreva, S, Lavrik, O (1992) Determination of tRNA(Phe) recognition nucleotides for phenylalanyl-tRNA synthetase from Thermus thermophilus. *Biochimie* 74:353–356.

[260] Moor, NA, Ankilova, VN, Lavrik, OI (1995) Recognition of tRNAPhe by phenylalanyl-tRNA synthetase of Thermus thermophilus. *Eur. J. Biochem.* 234:897–902.

[261] Vasil'eva, IA, Ankilova, VN, Lavrik, OI, Moor, NA (2002) tRNA discrimination by T. thermophilus phenylalanyl-tRNA synthetase at the binding step. *J. Mol. Recognit.* 15:188–196.

[262] McClain, WH, Schneider, J, Gabriel, K (1994) Distinctive acceptor-end structure and other determinants of Escherichia coli tRNAPro identity. *Nucleic Acids Res.* 22:522–529.

[263] Stehlin, C, Burke, B, Yang, F, Liu, H, Shiba, K, Musier-Forsyth, K (1998) Species-specific differences in the operational RNA code for aminoacylation of tRNAPro. *Biochemistry* 37:8605–8613.

[264] Burke, B, Yang, F, Chen, F, Stehlin, C, Chan, B, Musier-Forsyth, K (2000) Evolutionary coadaptation of the motif 2–acceptor stem interaction in the class II prolyl-tRNA synthetase system. *Biochemistry* 39:15540–15547.

[265] Hasegawa, T, Yokogawa, T (2000) Escherichia coli proline tRNA: structure and recognition sites for prolyl-tRNA synthetase. *Nucleic Acids Symp. Ser.* pp 7–8.

[266] Lipman, RS, Beuning, PJ, Musier-Forsyth, K, Hou, YM (2002) Amino acid activation of a dual-specificity tRNA synthetase is independent of tRNA. *J. Mol. Biol.* 316:421–427.

[267] Yokozawa, J, Okamoto, K, Kawarabayasi, Y, Kuno, A, Hasegawa, T (2003) Molecular recognition of proline tRNA by prolyl-tRNA synthetase from hyperthermophilic archaeon, Aeropyrum pernix K1. *Nucleic Acids Res. Suppl.* pp 247–248.

[268] Himeno, H, Hasegawa, T, Ueda, T, Watanabe, K, Shimizu, M (1990) Conversion of aminoacylation specificity from tRNA(Tyr) to tRNA(Ser) in vitro. *Nucleic Acids Res.* 18:6815–6819.

[269] Sampson, JR, Saks, ME (1993) Contributions of discrete tRNA(Ser) domains to aminoacylation by E.coli seryl-tRNA synthetase: a kinetic analysis using model RNA substrates. *Nucleic Acids Res.* 21:4467–4475.

[270] Asahara, H, Himeno, H, Tamura, K, Nameki, N, Hasegawa, T, Shimizu, M (1994) Escherichia coli seryl-tRNA synthetase recognizes tRNA(Ser) by its characteristic tertiary structure. *J. Mol. Biol.* 236:738–748.

[271] Saks, ME, Sampson, JR (1996) Variant minihelix RNAs reveal sequence-specific recognition of the helical tRNA(Ser) acceptor stem by E.coli seryl-tRNA synthetase. *EMBO J.* 15:2843–2849.

[272] Himeno, H, Yoshida, S, Soma, A, Nishikawa, K (1997) Only one nucleotide insertion to the long variable arm confers an efficient serine acceptor activity upon Saccharomyces cerevisiae tRNA(Leu) in vitro. *J. Mol. Biol.* 268:704–711.

[273] Achsel, T, Gross, HJ (1993) Identity determinants of human tRNA(Ser): sequence elements necessary for serylation and maturation of a tRNA with a long extra arm. *EMBO J.* 12:3333–3338.

[274] Lenhard, B, Orellana, O, Ibba, M, Weygand-Durasevi?, I (1999) tRNA recognition and evolution of determinants in seryl-tRNA synthesis. *Nucleic Acids Res.* 27:721–729.

[275] Gruic-Sovulj, I, Jaric, J, Dulic, M, Cindric, M, Weygand-Durasevic, I (2006) Shuffling of discrete tRNASer regions reveals differently utilized identity elements in yeast and methanogenic archaea. *J. Mol. Biol.* 361:128–139.

[276] Heckl, M, Busch, K, Gross, HJ (1998) Minimal tRNA(Ser) and tRNA(Sec) substrates for human seryl-tRNA synthetase: contribution of tRNA domains to serylation and tertiary structure. *FEBS Lett.* 427:315–319.

[277] Korencic, D, Polycarpo, C, Weygand-Durasevic, I, Söll, D (2004) Differential modes of transfer RNASer recognition in Methanosarcina barkeri. *J. Biol. Chem.* 279:48780–48786.

[278] Biou, V, Yaremchuk, A, Tukalo, M, Cusack, S (1994) The 2.9 A crystal structure of T. thermophilus seryl-tRNA synthetase complexed with tRNA(Ser). *Science* 263:1404–1410.

[279] Gomes, AC, Costa, T, Carreto, L, Santos, MA (2006) [The molecular mechanism of evolution of changes in the genetic code]. *Mol. Biol. (Mosk.)* 40:634–639.

[280] Korencic, D, Ahel, I, Söll, D (2002) Aminoacyl-tRNA Synthesis in Methanogenic Archaea. *Food Technol. Biotechnol.* 40:255–260.

[281] Itoh, Y, Sekine, S, Kuroishi, C, Terada, T, Shirouzu, M, Kuramitsu, S, Yokoyama, S (2008) Crystallographic and mutational studies of seryl-tRNA synthetase from the archaeon Pyrococcus horikoshii. *RNA Biol* 5:169–177.

[282] Schulman, LH, Pelka, H (1990) An anticodon change switches the identity of E. coli tRNA(mMet) from methionine to threonine. *Nucleic Acids Res.* 18:285–289.

[283] Hasegawa, T, Miyano, M, Himeno, H, Sano, Y, Kimura, K, Shimizu, M (1992) Identity determinants of E. coli threonine tRNA. *Biochem. Biophys. Res. Commun.* 184:478–484.

[284] Nameki, N (1995) Identity elements of tRNA(Thr) towards Saccharomyces cerevisiae threonyl-tRNA synthetase. *Nucleic Acids Res.* 23:2831–2836.

[285] Nameki, N, Asahara, H, Hasegawa, T (1996) Identity elements of Thermus thermophilus tRNA(Thr). *FEBS Lett.* 396:201–207.

[286] Nagaoka, Y, Yokozawa, J, Umehara, T, Iwaki, J, Okamoto, K, Kawarabayasi, Y, Koyama, Y, Sako, Y, Wakagi, T, Kuno, A, Hasegawa, T (2002) Molecular recognition of threonine tRNA by threonyl-tRNA synthetase from an extreme thermophilic archaeon, Aeropyrum pernix K1. *Nucleic Acids Res. Suppl.* pp 81–82.

[287] Ishikura, H, Nagaoka, Y, Yokozawa, J, Umehara, T, Kuno, A, Hasegawa, T (2000) Threonyl-tRNA synthetase of archaea: importance of the discriminator base in the aminoacylation of threonine tRNA. *Nucleic Acids Symp. Ser.* pp 83–84.

[288] Himeno, H, Hasegawa, T, Asahara, H, Tamura, K, Shimizu, M (1991) Identity determinants of E. coli tryptophan tRNA. *Nucleic Acids Res.* 19:6379–6382.

[289] Pak, M, Pallanck, L, Schulman, LH (1992) Conversion of a methionine initiator tRNA into a tryptophan-inserting elongator tRNA in vivo. *Biochemistry* 31:3303–3309.

[290] Rogers, MJ, Adachi, T, Inokuchi, H, Söll, D (1992) Switching tRNA(Gln) identity from glutamine to tryptophan. *Proc. Natl. Acad. Sci. U.S.A.* 89:3463–3467.

[291] Pak, M, Willis, IM, Schulman, LH (1994) Analysis of acceptor stem base pairing on tRNA(Trp) aminoacylation and function in vivo. *J. Biol. Chem.* 269:2277–2282.

[292] Yesland, KD, Johnson, JD (1993) Anticodon bases C34 and C35 are major, positive, identity elements in Saccharomyces cerevisiae tRNA(Trp). *Nucleic Acids Res.* 21:5079–5084.

[293] Xue, H, Shen, W, Giege, R, Wong, JT (1993) Identity elements of tRNA(Trp). Identification and evolutionary conservation. *J. Biol. Chem.* 268:9316–9322.

[294] Xu, F, Jiang, G, Li, W, He, X, Jin, Y, Wang, D (2002) Three GC base pairs Required for the Efficient Aminoacylation of tRNATrp by Tryptophanyl-tRNA Synthetase from Bacillus subtilis. *Biochemistry* 41:8087–8092.

[295] Guo, Q, Gong, Q, Tong, KL, Vestergaard, B, Costa, A, Desgres, J, Wong, M, Grosjean, H, Zhu, G, Wong, JT, Xue, H (2002) Recognition by tryptophanyl-tRNA synthetases of discriminator base on tRNATrp from three biological domains. *J. Biol. Chem.* 277:14343–14349.

[296] Ulmasov, B, Topin, A, Chen, Z, He, SH, Folk, WR (1998) Identity elements and aminoacylation of plant tRNATrp. *Nucleic Acids Res.* 26:5139–5141.

[297] Xu, F, Chen, X, Li, X, Chen, L, Jin, Y, Wang, D (2001) Species-specific differences in the operational RNA code for aminoacylation of tRNATrp. *Nucl. Acids Res.* 29:4125–4133.

[298] Shen, N, Guo, L, Yang, B, Jin, Y, Ding, J (2006) Structure of human tryptophanyl-tRNA synthetase in complex with tRNATrp reveals the molecular basis of tRNA recognition and specificity . *Nucl. Acids Res.* 34:3246–3258.

[299] Gong, Q, Guo, Q, Tong, KL, Zhu, G, Wong, JT, Xue, H (2002) NMR analysis of bovine tRNATrp: conformation dependence of Mg2+ binding. *J. Biol. Chem.* 277:20694–20701.

[300] Bedouelle, H (1990) Recognition of tRNA(Tyr) by tyrosyl-tRNA synthetase. *Biochimie* 72:589–598.

[301] Sherman, JM, Rogers, K, Rogers, MJ, Söll, D (1992) Synthetase competition and tRNA context determine the in vivo identify of tRNA discriminator mutants. *J. Mol. Biol.* 228:1055–1062.

[302] Bare, LA, Uhlenbeck, OC (1986) Specific substitution into the anticodon loop of yeast tyrosine transfer RNA. *Biochemistry* 25:5825–5830.

[303] Quinn, CL, Tao, N, Schimmel, P (1995) Species-specific microhelix aminoacylation by a eukaryotic pathogen tRNA synthetase dependent on a single base pair. *Biochemistry* 34:12489–12495.

[304] Iwaki, J, Asahara, H, Nagaoka, Y, Yokozawa, J, Umehara, T, Kawarabayasi, Y, Koyama, Y, Sako, Y, Kuno, A, Hasegawa, T (2002) Differences in tyrosine tRNA identity between Escherichia coli and archaeon, Aeropyrum pernix K1. *Nucleic Acids Res. Suppl.* pp 225–226.

[305] Bonnefond, L, Gieg, R, Rudinger-Thirion, J (2005) Evolution of the tRNA(Tyr)/TyrRS aminoacylation systems. *Biochimie* 87:873–883.

[306] Fechter, P, Rudinger-Thirion, J, Theobald-Dietrich, A, Giege, R (2000) Identity of tRNA for yeast tyrosyl-tRNA synthetase: tyrosylation is more sensitive to identity nucleotides than to structural features. *Biochemistry* 39:1725–1733.

[307] Fechter, P, Rudinger-Thirion, J, Tukalo, M, Giege, R (2001) Major tyrosine identity determinants in Methanococcus jannaschii and Saccharomyces cerevisiae tRNA(Tyr) are conserved but expressed differently. *Eur. J. Biochem.* 268:761–767.

[308] Wakasugi, K, Quinn, CL, Tao, N, Schimmel, P (1998) Genetic code in evolution: switching species-specific aminoacylation with a peptide transplant. *EMBO J.* 17:297–305.

[309] Kisselev, LL (1985) The role of the anticodon in recognition of tRNA by aminoacyl-tRNA synthetases. *Prog. Nucleic Acid Res. Mol. Biol.* 32:237–266.

[310] Chu, WC, Horowitz, J (1991) Recognition of Escherichia coli valine transfer RNA by its cognate synthetase: a fluorine-19 NMR study. *Biochemistry* 30:1655–1663.

[311] Tamura, K, Himeno, H, Asahara, H, Hasegawa, T, Shimizu, M (1991) Identity determinants of E. coli tRNA(Val). *Biochem. Biophys. Res. Commun.* 177:619–623.

[312] Kern, D, Giege, R, Ebel, JP (1972) Incorrect aminoacylatins catalysed by the phenylalanyl-and valyl-tRNA synthetases from yeast. *Eur. J. Biochem.* 31:148–155.

[313] Florentz, C, Dreher, TW, Rudinger, J, Giege, R (1991) Specific valylation identity of turnip yellow mosaic virus RNA by yeast valyl-tRNA synthetase is directed by the anticodon in a kinetic rather than affinity-based discrimination. *Eur. J. Biochem.* 195:229–234.

[314] Fukai, S, Nureki, O, Sekine, S, Shimada, A, Vassylyev, DG, Yokoyama, S (2003) Mechanism of molecular interactions for tRNA(Val) recognition by valyl-tRNA synthetase. *RNA* 9:100–111.

[315] Yang, XL, Otero, FJ, Ewalt, KL, Liu, J, Swairjo, MA, Kohrer, C, RajBhandary, UL, Skene, RJ, McRee, DE, Schimmel, P (2006) Two conformations of a crystalline human tRNA synthetase-tRNA complex: implications for protein synthesis. *EMBO J.* 25:2919–2929.

[316] Sekine, SI, Nureki, O, Dubois, DY, Bernier, S, Chenevert, R, Lapointe, J, Vassylyev, DG, Yokoyama, S (2003) ATP binding by glutamyl-tRNA synthetase is switched to the productive mode by tRNA binding. *EMBO J.* 22:676–688.

[317] Perona, JJ, Rould, MA, Steitz, TA (1993) Structural basis for transfer RNA aminoacylation by Escherichia coli glutaminyl-tRNA synthetase. *Biochemistry* 32:8758–8771.

[318] Eriani, G, Delarue, M, Poch, O, Gangloff, J, Moras, D (1990) Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature* 347:203–206.

[319] Cusack, S (1995) Eleven down and nine to go. *Nature Struct. Biol.* 2:824–831.

[320] Landes, C, Perona, JJ, Brunie, S, Rould, MA, Zelwer, C, Steitz, TA, Risler, JL (1995) A structure-based multiple sequence alignment of all class I aminoacyl-tRNA synthetases. *Biochimie* 77:194–203.

[321] O'Donoghue, P, Luthey-Schulten, Z (2005) Evolutionary profiles derived from the qr factorization of multiple strucutral alignments gives an economy of information. *J. Mol. Biol.* 346:875–894.

[322] Nureki, O, Vassylyev, D, Katayanagi, K, Shimizu, T, Sekine, S, Kigawa, T, Miyazawa, T, Yokoyama, S, Morikawa, K (1995) Architectures of class-defining and specific domains of glutamyl-tRNA synthetase. *Science* 267:1958–1965.

[323] Tateno, M, Nureki, O, Sekine, S, Kaneda, K, Go, M, Yokoyama, S (1995) A three-dimensional structure model of the complex of glutamyl-tRNA synthetase and its cognate tRNA. *FEBS Lett.* 377:77–81.

[324] Sekine, S, Nureki, O, Shimada, A, Vassylyev, DG, Yokoyama, S (2001) Structural basis for anticodon recognition by discriminating glutamyl-tRNA synthetase. *Nature Struct. Biol.* 8:203–206.

[325] Siatecka, M, Rozek, M, Barciszewski, J, Mirande, M (1998) Modular evolution of the Glx-tRNA synthetase family–rooting of the evolutionary tree between the bacteria and archaea/eukarya branches. *Eur. J. Biochem.* 256:80–87.

[326] Dubois, DY, Blais, SP, Huot, JL, Lapointe, J (2009) A C-truncated glutamyl-tRNA synthetase specific for tRNA(Glu) is stimulated by its free complementary distal domain: mechanistic and evolutionary implications. *Biochemistry* 48:6012–6021.

[327] Lee, J, Hendrickson, T (2004) Divergent anticodon recognition in contrasting glutamyl-tRNA synthetases. *J. Mol. Biol.* 344:1167–1174.

[328] Lamour, V, Quevillon, S, Diriong, S, N'Guyen, VC, Lipinski, M, Mirande, M (1994) Evolution of the Glx-tRNA synthetase family: the glutaminyl enzyme as a case of horizontal gene transfer. *Proc. Natl. Acad. Sci. USA* 91:8670–8674.

[329] Salazar, JC, Ahel, I, Orellana, O, Tumbula-Hansen, D, Krieger, R, Daniels, L, Sll, D (2003) Coevolution of an aminoacyl-tRNA synthetase with its tRNA substrates. *Proc. Natl. Acad. Sci. U.S.A.* 100:13863–13868.

[330] Skouloubris, S, Ribas de Pouplana, L, De Reuse, H, Hendrickson, T (2003) A noncognate aminoacyl-tRNA synthetase that may resolve a missing link in protein evolution. *Proc. Natl. Acad. Sci. USA* 100:11297–11302.

[331] Budiman, ME, Knaggs, MH, Fetrow, JS, Alexander, RW (2007) Using molecular dynamics to map interaction networks in an aminoacyl-tRNA synthetase. *PROTEINS: Structure, Function, and Genetics* 68:670–689.

[332] Hansia, P, Ghosh, A, Vishveshwara, S (2009) Ligand dependent intra and inter subunit communication in human tryptophanyl tRNA synthetase as deduced from the dynamics of structure networks. *Mol Biosyst* 5:1860–1872.

[333] Bharatham, N, Bharatham, K, Lee, Y, Woo Lee, K (2009) Molecular dynamics simulation study of valyl-tRNA synthetase with its pre- and post-transfer editing substrates. *Biophys. Chem.* 143:34–43.

[334] Archontis, G, Simonson, T, Karplus, M (2001) Binding free energies and free energy components from molecular dynamics and Poisson-Boltzmann calculations. Application to amino acid recognition by aspartyl-tRNA synthetase. *J. Mol. Biol.* 306:307–327.

[335] Hughes, SJ, Tanner, JA, Hindley, AD, Miller, AD, Gould, IR (2003) Functional asymmetry in the lysyl-tRNA synthetase explored by molecular dynamics, free energy calculations and experiment. *BMC Structural Biology* 3.

[336] Kapustina, M, Carter, C (2006) Computational studies of tryptophanyl-tRNA synthetase: activation of ATP by induced-fit. *J. Mol. Biol.* 362:1159–1180.

[337] Yamasaki, S, Nakamura, S, Terada, T, Shimizu, K (2007) Mechanism of the difference in the binding affinity of E. coli $tRNA^{Gln}$ to glutaminyl-tRNA synthetase caused by noninterface nucleotides in variable loop. *Biophys. J.* 92:192–200.

[338] Thompson, D, Lazennec, C, Plateau, P, Simonson, T (2008) Probing electrostatic interactions and ligand binding in aspartyl-tRNA synthetase through site-directed mutagenesis and computer simulations. *PROTEINS: Structure, Function, and Genetics* 71:1450–1460.

[339] McGinnis, S, Madden, TL (2004) BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucl. Acids Res.* 32:W20–W25.

[340] Wheeler, DL, Barrett, T, Benson, DA, Bryant, SH, Canese, K, Chetvernin, V, Church, DM, DiCuccio, M, Edgar, R, Federhen, S, Geer, LY, Helmberg, W, Kapustin, Y, Kenton, DL, Khovayko, O, Lipman, DJ, Madden, TL, Maglott, DR, Ostell, J, Pruitt, KD, Schuler, GD, Schriml, LM, Sequeira, E, Sherry, ST, Sirotkin, K, Souvorov, A, Starchenko, G, Suzek, TO, Tatusov, R, Tatusova, TA, Wagner, L, Yaschenko, E (2006) Database resources of the National Center for Biotechnology Information. *Nucl. Acids Res.* 34:173–180.

[341] Sprinzl, M, Horn, C, Brown, M, Ioudovitch, A, Steinberg, S (1998) Compilation of tRNA sequences and sequences of tRNA genes. *Nucl. Acids Res.* 26:148–153.

[342] Rath, VL, Silvian, LF, Beijer, B, Sproat, BS, Steitz, TA (1998) How glutaminyl-tRNA synthetase selects glutamine. *Structure* 6:439–449.

[343] Delagoutte, B, Moras, D, Cavarelli, J (2000) tRNA aminoacylation by arginyl-tRNA synthetase: induced conformations during substrates binding. *EMBO J.* 19:5599–5610.

[344] Newberry, KJ, Hou, YM, Perona, JJ (2002) Structural origins of amino acid selection without editing by cysteinyl-tRNA synthetase. *EMBO J.* 21:2778–2787.

[345] Silvian, LF, Wang, J, Steitz, TA (1999) Insights into editing from an ile-tRNA synthetase structure with tRNAile and mupirocin. *Science* 285:1074–1077.

[346] Cusack, S, Yaremchuk, A, Tukalo, M (2000) The 2 Å crystal structure of leucyl-tRNA synthetase and its complex with a leucyl-adenylate analogue. *EMBO J.* 19:2351–2361.

[347] Nakanishi, K, Ogiso, Y, Nakama, T, Fukai, S, Nureki, O (2005) Structural basis for anticodon recognition by methionyl-tRNA synthetase. *Nat. Struct. Mol. Biol.* 12:931–932.

[348] Fukai, S, Nureki, O, Sekine, S, Shimada, A, Tao, J, Vassylyev, DG, Yokoyama, S (2000) Structural basis for double-sieve discrimination of L-valine from L-isoleucine and L-threonine by the complex of tRNA(Val) and valyl-tRNA synthetase. *Cell* 103:793–803.

[349] Yaremchuk, A, Kriklivyi, I, Tukalo, M, Cusack, S (2002) Class I tyrosyl-tRNA synthetase has a class II mode of cognate tRNA recognition. *EMBO J.* 21:3829–3840.

[350] Bairoch, A, Apweiler, R, Wu, CH, Barker, WC, Boeckmann, B, Ferro, S, Gasteiger, E, Huang, H, Lopez, R, Magrane, M, Martin, MJ, Natale, DA, O'Donovan, C, Redaschi, N, Yeh, LSL (2005) The Universal Protein Resource (UniProt). *Nucl. Acids Res.* 33:154–159.

[351] Vriend, G (1990) WHAT IF: a molecular modeling and drug design program. *J. Mol. Graphics* 8:52–56.

[352] Jorgensen, W, Chandrasekhar, J, Madura, J, Impey, R, Klein, M (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.

[353] Draper, DE (2004) A guide to ions and RNA structure. *RNA* 10:335–343.

[354] Phillips, JC, Braun, R, Wang, W, Gumbart, J, Tajkhorshid, E, Villa, E, Chipot, C, Skeel, RD, Kale, L, Schulten, K (2005) Scalable molecular dynamics with NAMD. *J. Comp. Chem.* 26:1781–1802.

[355] Foloppe, N, MacKerrell Jr., AD (2000) All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comp. Chem.* 21:86–104.

[356] Feller, S, Zhang, Y, Pastor, R, Brooks, B (1995) Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* 103:4613–4621.

[357] Darden, T, York, D, Pedersen, L (1993) Particle mesh Ewald: An $N\cdot\log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* 89:10089–10092.

[358] Auffinger, P, Westhof, E (1997) RNA hydration: three nanoseconds of multiple molecular dynamics simulations of the solvated tRNA(Asp) anticodon hairpin. *J. Mol. Biol.* 269:326–341.

[359] Kabsch, W (1978) A discussion of the solution for the best rotation to relate two sets of vectors. *Acta Cryst.* A34:827–828.

[360] Bas, DC, Rogers, DM, Jensen, JH (2008) Very fast prediction and rationalization of pKa values for protein-ligand complexes. *PROTEINS: Structure, Function, and Genetics* 73:765–783.

[361] Glykos, NM (2006) Software news and updates. Carma: a molecular dynamics analysis program. *J. Comp. Chem.* 27:1765–1768.

[362] Floyd, RW (1962) Algorithm 97: Shortest Path. *Communications of the ACM* 5:345.

[363] (2001) *Introduction to Algorithms, 2nd edition* (MIT Press, Cambridge, MA).

[364] Girvan, M, Newman, M (2002) Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* 99:7821–7826.

[365] Froloff, N, Windemuth, A, Honig, B (1997) On the calculation of binding free energies using continuum methods: application to MHC class I protein-peptide interactions. *Protein Science* 6:1293–1301.

[366] Kollman, PA, Massova, I, Reyes, C, Kuhn, B, Huo, S, Chong, L, Lee, M, Lee, T, Duan, Y, Wang, W, Donini, O, Cieplak, P, Srinivasan, J, Case, DA, Cheatham 3rd, TE (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* 33:889–897.

[367] Reyes, CM, Kollman, PA (2000) Structure and thermodynamics of RNA-protein binding: using molecular dynamics and free energy analyses to calculate the free energies of binding and conformational change. *J. Mol. Biol.* 297:1145–1158.

[368] Rocchia, W, Alexov, E, Honig, B (2001) Extending the Applicability of the Nonlinear Poisson-Boltzmann Equation: Multiple Dielectric Constants and Multivalent Ions. *J. Phys. Chem. B* 105:6507–6514.

[369] Gohlke, H, Kiel, C, Case, DA (2003) Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes. *J. Mol. Biol.* 330:891–913.

[370] Pogorelov, TV, Autenrieth, F, Roberts, E, Luthey-Schulten, ZA (2007) Cytochrome c(2) Exit Strategy: Dissociation Studies and Evolutionary Implications. *J Phys Chem B* 111:618–634.

[371] Wong, S, Amaro, R, McCammon, J (2009) MM-PBSA captures a key role of intercalating water molecules at a protein:protein interface. *J. Chem. Theory Comp.* 5:422–429.

[372] Baker, NA, Sept, D, Joseph, S, Holst, MJ, McCammon, JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* 98:10037–10041.

[373] Schlitter, J (1993) Estimation of absolute and relative entropies of macromolecules using the covariance matrix. *Chem. Phys. Lett.* 215:617–621.

[374] Schäfer, A, Mark, AE, van Gunsteren, WF (2000) Absolute entropies from molecular dynamics simulation trajectories. *J. Chem. Phys.* 113:7809–7817.

[375] Anricrioaei, I, Karplus, M (2001) On the calculation of entropy from covariance matrices of the atomic fluctuations. *J. Chem. Phys.* 115:6289–6292.

[376] Gupta, MC (1991) *Statistical Thermodynamics* (Halsted Press).

[377] Xin, Y, Li, W, First, E (2000) Stabilization of the Transition State for the Transfer of Tyrosine to $tRNA^{Tyr}$ by Tyrosyl-tRNA Synthetase. *JMB* 303:299–310.

[378] Ganguly, S, Kundu, KK (1994) Deprotonation energetics of adenine, 5′–adenosine monophosphate and adenosine triphosphate in water from emf and spectrophotometic measurements. *J. Sol. Chem.* 23:1227–1246.

[379] Bishop, AC, Nomanbhoy, TK, Schimmel, P (2002) Blocking site-to-site translocation of a misactivated amino acid by mutation of a class I tRNA synthetase. *Proc. Natl. Acad. Sci. U.S.A.* 99:585–590.

[380] Lincecum, TL, Tukalo, M, Yaremchuk, A, Mursinna, RS, Williams, AM, Sproat, BS, Van Den Eynde, W, Link, A, Van Calenbergh, S, Grøtli, M, Martinis, SA, Cusack, S (2003) Structural and mechanistic basis of pre- and posttransfer editing by leucyl-tRNA synthetase. *Mol. Cell* 11:951–963.

[381] Tsunoda, M, Kusakabe, Y, Tanaka, N, Ohno, S, Nakamura, M, Senda, T, Moriguchi, T, Asai, N, Sekine, M, Yokogawa, T, Nishikawa, K, Nakamura, KT (2007) Structural basis for recognition of cognate tRNA by tyrosyl-tRNA synthetase from three kingdoms. *Nucl. Acids Res.* 35:4289–4300.

[382] Sekine, S, Nureki, O, Tateno, M, Yokoyama, S (1999) The identity determinants required for the discrimination between $tRNA^{Glu}$ and $tRNA^{Asp}$ by glutamyl-tRNA synthetase from Escherichia coli. *Eur. J. Biochem.* 261:354.

[383] Hong, K, Ibba, M, Weygand-Durasevic, I, Rogers, M, Thomann, H, Söll, D (1996) Transfer RNA-dependent cognate amino acid recognition by an aminoacyl-tRNA synthetase. *EMBO J.* 15:1983–1991.

[384] O'Donoghue, P, Sethi, A, Woese, CR, Luthey-Schulten, ZA (2005) The evolutionary history of cys-trna(cys) formation. *Proc. Natl. Acad. Sci. USA* 102:19003–19008.

[385] Zhang, CM, Perona, JJ, Hou, YM (2003) Amino acid discrimination by a highly differentiated metal center of an aminoacyl-tRNA synthetase. *Biochemistry* 42:10931–10937.

[386] Zhang, C, Liu, C, Slater, S, Hou, Y (2008) Aminoacylation of tRNA with phosphoserine for synthesis of cysteinyl-tRNA(Cys). *Nat. Struct. Mol. Biol.* 15:507–514.

[387] Roberts, E, Montoya, J, Rosenfeld, E, Luthey-Schulten, Z (2009) Unbiased evolutionary profiles created using QR at maximum Shannon entropy. In preparation.

[388] Fukunaga, R, Yokoyama, S (2005) Aminoacylation complex structures of leucyl-tRNA synthetase and tRNALeu reveal two modes of discriminator-base recognition. *Nat. Struct. Mol. Biol.* 12:915–922.

[389] Lo Conte, L, Ailey, B, Hubbard, TJ, Brenner, SE, Murzin, AG, Chothia, C (2000) SCOP: a structural classification of proteins database. *Nucl. Acids Res.* 28:257–259.

[390] Marti-Renom, MA, Stuart, AC, Fiser, A, Sanchez, R, Melo, F, Sali, A (2000) Comparative protein structure modeling of genes and genomes. *Ann. Rev. Biophys. Biomol. Struct.* 29:291–325.

[391] Stone, JE, Gullingsrud, J, Schulten, K, Grayson, P (2001) A System for Interactive Molecular Dynamics Simulation. *ACM Symposium on Interactive 3D Graphics* pp 191–194.

[392] Simonson, T, Calimet, N (2002) CysxHisy–Zn2+ Interactions: Thiol vs. Thiolate Coordinate. *PROTEINS: Structure, Function, and Genetics* 49:37–48.

[393] Nilsson, L, Karplus, M (1986) Empirical energy functions for energy minimization and dynamics of nucleic acids. *J. Comp. Chem.* 7:591–616.

[394] Schmidt, MW, Baldridge, KK, Boatz, JA, Elbert, ST, Gordon, MS, Jensen, JH, Koseki, S, Matsunaga, N, Nguyen, KA, Su, S, Windus, TL, Dupuis, M, Montgomery, JA (1993) General atomic and molecular electronic structure system. *J. Comp. Chem.* 14:1347–1363.

[395] Nakamura, Y, Gojobori, T, Ikemura, T (2000) Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucl. Acids Res.* 28:292.

[396] Shimizu, S, Juan, EC, Sato, Y, Miyashita, Y, Hoque, MM, Suzuki, K, Sagara, T, Tsunoda, M, Sekiguchi, T, Dock-Bregeon, AC, Moras, D, Taknaka, A (2009) Two complementary enzymes for threonylation of tRNA in crenarchaeota: crystal structure of Aeropyrum pernix threonyl-tRNA synthetase lacking a cis-editing domain. *J. Mol. Biol.* 394:286–296.

[397] Sankaranarayanan, R, Dock-Bregeon, AC, Romby, P, Caillet, J, Springer, M, Rees, B, Ehresmann, C, Ehresmann, B, Moras, D (1999) The structure of threonyl-tRNA synthetase-tRNA(Thr) complex enlightens its repressor activity and reveals an essential zinc ion in the active site. *Cell* 97:371–381.

[398] Hamann, CS, Hou, YM (1997) An RNA structural determinant for tRNA recognition. *Biochemistry* 36:7967–7972.

[399] Hamann, CS, Hou, YM (2000) Probing a tRNA core that contributes to aminoacylation. *J. Mol. Biol.* 295:777–789.

[400] Hohn, M, Park, H, O'Donoghue, P, Schnitzbauer, M, Söll, D (2006) Emergence of the universal genetic code imprinted in an RNA record. *Proc. Natl. Acad. Sci. USA* 103:18095–18100.

[401] Konno, M, Sumida, T, Uchikawa, E, Mori, Y, Yanagisawa, T, Sekine, S, Yokoyama, S, Yokoyama, S (2009) Modeling of tRNA-assisted mechanism of Arg activation based on a structure of Arg-tRNA synthetase, tRNA, and an ATP analog (ANP). *FEBS J.* 276:4763–4779.

[402] Li, S, Pelka, H, Schulman, LH (1993) The anticodon and discriminator base are important for aminoacylation of Escherichia coli tRNA(Asn). *J. Biol. Chem.* 268:18335–18339.

[403] Bailly, M, Giannouli, S, Blaise, M, Stathopoulos, C, Kern, D, Becker, HD (2006) A single tRNA base pair mediates bacterial tRNA-dependent biosynthesis of asparagine. *Nucleic Acids Res.* 34:6083–6094.

[404] Eiler, S, Dock-Bregeon, A, Moulinier, L, Thierry, JC, Moras, D (1999) Synthesis of aspartyl-tRNA(Asp) in Escherichia coli–a snapshot of the second step. *EMBO J.* 18:6532–6541.

[405] Ruff, M, Krishnaswamy, S, Boeglin, M, Poterszman, A, Mitschler, A, Podjarny, A, Rees, B, Thierry, JC, Moras, D (1991) Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with tRNA(Asp). *Science* 252:1682–1689.

[406] Nakanishi, K, Ogiso, Y, Nakama, T, Fukai, S, Nureki, O (2005) Structural basis for anticodon recognition by methionyl-tRNA synthetase. *Nat. Struct. Mol. Biol.* 12:931–932.

[407] Moor, N, Kotik-Kogan, O, Tworowski, D, Sukhanova, M, Safro, M (2006) The crystal structure of the ternary complex of phenylalanyl-tRNA synthetase with tRNAPhe and a phenylalanyl-adenylate analogue reveals a conformational switch of the CCA end. *Biochemistry* 45:10572–10583.

[408] Yaremchuk, A, Tukalo, M, Grotli, M, Cusack, S (2001) A succession of substrate induced conformational changes ensures the amino acid specificity of Thermus thermophilus prolyl-tRNA synthetase: comparison with histidyl-tRNA synthetase. *J. Mol. Biol.* 309:989–1002.

AUTHOR'S BIOGRAPHY

Alexis Black Pyrkosz was born in Dearborn, Michigan in 1982. She attended Lawrence Technological University in Southfield, Michigan and graduated Summa cum Laude with a Bachelor of Science, majoring in Chemistry and minoring in Computer Science in 2004. As an undergraduate, Alexis completed two Research Experience for Undergraduates programs: the first with Mark Hollingsworth at Kansas State University studying ferroelastic properties of urea inclusion crystals, and the second with Donald Burke at Indiana University at Bloomington performing kinetics studies of mutant hammerhead ribozymes.

Alexis came to the University of Illinois at Urbana-Champaign in 2004 and worked for Prof. Zaida A. Luthey-Schulten. Alexis's thesis work involved investigating the evolution and dynamical behavior of transfer RNA during the first two steps of translation. She received her Ph.D. in April 2010.